#6

Kongeriget Danmark

Patent application No.: PA 1999 00092

Date of filing: 22 JAN. 1999

Applicant: Novo Nordisk A/S, Novo Allé
DK-2880 Bagsværd

This is to certify the correctness of the following information:

The attached photocopy is a true copy of the following document:

- The specification, claims, figures and abstract as filed with the application on the filing date indicated above.

Patentdirektoratet · Kongeriget Danmark ·

Erhvervsministeriet
Patentdirektoratet

TAASTRUP 10 Sept. 1999

Karin Schlichting
Head Clerk

Phytases (*myo*-inositol hexakisphosphate phosphohydrolases; EC 3.1.3.8) are enzymes that hydrolyze phytate (*myo*-inositol hexakisphosphate) to *myo*-inositol and inorganic phosphate and are known to be valuable feed additives.

A phytase was first described in rice bran in 1907 [Suzuki et al., Bull. Coll. Agr. Tokio Imp. Univ. 7, 495 (1907)] and phytases from Aspergillus species in 1911 [Dox and Golden, J. Biol. Chem. 10, 183-186 (1911)]. Phytases have also been found in wheat bran, plant seeds, animal intestines and in microorganisms [Howsen and Davis, Enzyme Microb. Technol. 5, 377-382 (1983), Lambrechts et al., Biotech. Lett. 14, 61-66 (1992), Shieh and Ware, Appl. Microbiol. 16, 1348-1351 (1968)].

The cloning and expression of the phytase from Aspergillus niger (ficuum) has been described by Van Hartingsveldt et al., in Gene, 127, 87-94 (1993) and in European Patent Application, Publication No. (EP) 420 358 and from Aspergillus niger var. awamori by Piddington et al., in Gene 133, 55-62 (1993).

Cloning, expression and purification of phytases with improved properties have been disclosed in EP 684 313. However, since there is a still ongoing need for further improved phytases, especially with respect to their thermostability, it is an object of the present invention to provide the following process which is, however, not only applicable to phytases.

A process for the preparation of a consensus protein, whereby such process is characterized by the following steps:

a) at least three, preferably four amino acid sequences of a defined protein family are aligned by any standard alignment program known in the art;

b) amino acids at the same position according to such alignment are compared regarding their evolutionary similarity by any standard program known in the art, whereas the degree of similarity provided by such a program which defines the least similarity of the amino acids that is used for the determination of an amino acid of corresponding positions is set to a less stringent number and the parameters are set in such a way that it is possible for the program to determine

from only 2 identical amino acids at a corresponding position an amino acid for the consensus protein; however, if among the compared amino acid sequences are sequences that show a much higher degree of similarity to each other than to the residual sequences, these sequences are represented by their consensus sequence determined as defined in the same way as in the present process for the consensus sequence of the consensus protein or a vote weight of 1 divided by the number of such sequences is assigned to every of those sequences;

c) in case no common amino acid at a defined position can be identified by the program, any of the amino acids of all sequences used for the comparison, preferably the most frequent amino acid of all such sequences is selected or an amino acid is selected on the basis of the consideration given in Example 2;

d) once the consensus sequence has been defined, such sequence is back-translated into a DNA sequence, preferably using a codon frequency table of the organism in which expression should take place;

e) the DNA sequence is synthesized by methods known in the art and used either integrated into a suitable expression vector or by itself to transform an appropriate host cell;

f) the transformed host cell is grown under suitable culture conditions and the consensus protein is isolated from the host cell or its culture medium by methods known in the art.

In a preferred embodiment of this process step b) can also be defined as follows: b) amino acids at the same position according to such an alignment are compared regarding their evolutionary similarity by any standard program known in the art, whereas the degree of similarity provided by such program is set at the lowest possible value and the amino acid which is the most similar for at least half of the sequences used for the comparison is selected for the corresponding position in the amino acid sequence of the consensus protein.

In another preferred embodiment the consensus sequence is used in order to improve a specific protein. In this process first a consensus sequence is determined from a number of highly homologous sequences according to steps a), b) and c) as described above. In a second step the amino acid sequence of another protein which is homologous to the consensus sequence is compared with the consensus sequence and in a third step only those amino acid residues are replaced in the amino acid sequence of the other protein

which clearly differ from the consensus sequence of this protein family calculated under moderately stringent conditions whereas at all positions of the alignment where no preferred single amino acid can be determined under moderately stringent conditions the amino acids of the other protein remain unchanged.

By using this preferred embodiment the consensus sequence derived from a number of highly homologous sequences is used in order to replace only certain amino acid residues in the protein in such a manner that only those amino acid residues are replaced which clearly and unambiguously differ from the corresponding consensus sequence of this protein family which has been calculated on moderately stringent conditions. At all other positions of the alignment, however, where the method of the present invention is not able to determine clearly a preferred amino acid residue under moderately stringent conditions the amino acid residues of the other protein are maintained unchanged.

A further preferred embodiment is a process wherein at first a consensus sequence is determined from homologous sequences as described above. In a second step the active center of the protein comprising all amino acid residues that are involved in forming the active center is determined in the consensus sequence and in the sequence of a single homologous protein as well. The single homologous protein may have preferred properties like high specific activity or different pH dependency of enzymatic activity. In a third step some or all amino acid residues that are involved in forming the active centre of the homologeous protein are inserted into the backbone of the consensus sequence. The result thereof is a chimeric protein having the active centre derived from a single protein and the backbone of the consensus sequence.

The active centre of the protein can be determined e.g. by using any analysis of the three-dimensional structure of the protein, e.g. by homology modelling on the basis of a known 3D-structure of a known protein. Frequently the single homologeous protein is an enzyme.

It is furthermore an object of the present invention to provide such a process, wherein the program used for the comparison of amino acids at a defined position regarding their evolutionary similarity is the program "PRETTY". It is more specifically an object of the present invention to provide such a process, wherein the defined protein family is the family of phytases, especially wherein the phytases are of fungal origin.

It is furthermore an object of the present invention to provide such processes, wherein the host cell is of eukaryotic, especially fungal, preferably *Aspergillus* or yeast, preferably *Saccharomyces* or *Hansenula* origin.

It is also an object of the present invention to provide a consensus protein obtainable preferably obtained, by such processes and specifically the consensus protein, which has the amino acid *sequences shown in Figures 2, 4 and 6* or a variant thereof. A "variant" refers in the context of the present invention to a consensus protein with amino acid sequence shown in Figure 2, 5, 7, and 8 wherein at one or more positions amino acids have been deleted, added or replaced by one or more other amino acids with the proviso that the resulting sequence provides for a protein whose basic properties like enzymatic activity (type of and specific activity), thermostability, activity in a certain pH-range (pH-stability) have not significantly been changed. "Significantly" means in this context that a man skilled in the art would say that the properties of the variant may still be different but would not be unobvious over the ones of the consensus protein with the amino acid sequence of Figure 2 itself.

A "mutein" refers in the context of the present invention to replacements of the amino acid in the amino acid sequences of the consensus proteins shown in Figure 2 which lead to consensus proteins with further improved properties e.g. activity. Such muteins can be defined and prepared on the basis of the teachings given in European Patent Application number 97810175.6, e. g. Q50L, Q50T, Q50G, Q50L-Y51N, or Q50T-Y51N. "Q50L" means in this context that at position 50 of the amino acid sequence (Figure 2) the amino acid Q has been replaced by amino acid L.

In addition, a food, feed or pharmaceutical composition comprising a consensus protein as defined above is also an object of the present invention.

In this context "at least three preferably four amino acid sequences of such defined protein family" means that three, four, five, six to 12, 20, 50 or even more sequences can be used for the alignment and the comparison to create the amino acid sequence of the consensus protein. "Sequences of a defined protein family" means that such sequences fold into a three dimensional structure, wherein the alpha-helixes, the beta-sheets and beta-turns are at the same position so that such structures are, as called by the man skilled in the art, largely superimposable. Furthermore these sequences characterize proteins which show the same type of biological activity, e.g. a defined enzyme class, e.g. the phytases. As known in the art, the three dimensional structure of one of such sequences is sufficient to allow the modelling of the structure of the other sequences of such a family. An example, how this can be effected, is given in the Reference Example of the present case. "Evolutionary similarity" in the context of the present invention refers to a scheme which classifies amino acids regarding their structural similarity which allows that one amino acid can be replaced by another amino acid with a minimal influence on the overall structure, as this is done e.g.

by programs, like "PRETTY", known in the art. The phrase "the degree of similarity provided by such a program...is set to less stringent number" means in the context of the present invention that values for the parameters which determine the degree of similarity in the program used in the practice of the present invention are chosen in a way to allow the program to define a common amino acid for a maximum of positions of the whole amino acid sequence, e. g. in case of the program PRETTY a value of 2 or 3 for the THRESHOLD and a value of 2 for the PLURALITY can be choosen. Furthermore, "a vote weight of one divided by the number of such sequences" means in the context of the present invention that the sequences which define a group of sequences with a higher degree of similarity as the other sequences used for the determination of the consensus sequence only contribute to such determination with a factor which is equal to one divided by a number of all sequences of this group.

As mentioned before should the program not allow to select the most similar amino acid, the most frequent amino acid is selected, should the latter be impossible the man skilled in the art will select an amino acid from all the sequences used for the comparison which is known in the art for its property to improve the thermostability in proteins as discussed e.g. by

Janecek, S. (1993), *Process Biochem.* 28, 435-445 or

Fersht, A. R. & Serrano, L. (1993), *Curr. Opin. Struct. Biol.* 3, 75-83.

Alber, T. (1989), *Annu. Rev. Biochem.* 58, 765-798 or

Matthews, B. W. (1987), *Biochemistry* 26, 6885-6888.

Matthews, B. W. (1991), *Curr. Opin. Struct. Biol.* **1**, 17-21.

The stability of an enzyme is a critical factor for many industrial applications. Therefore, a lot of attempts, more or less successful, have been made to improve the stability, preferably the thermostability of enzymes by rational (van den Burg *et al.*, 1998) or irrational approaches (Akanuma *et al.*, 1998). The forces influencing the thermostability of a protein are the same as those that are responsible for the proper folding of a peptide strand (hydrophobic interactions, van der Waals interactions, H-bonds, salt bridges, conformational strain (Matthews, 1993). Furthermore, as shown by Matthews *et al.* (1987), the free energy of the unfolded state has also an influence on the stability of a protein. Enhancing of protein stability means to increase the number and strength of favorable interactions and to decrease the number and strength of unfavorable interactions. It has been possible to introduce disulfide linkages (Sauer et al, 1986) to replace glycine with

alanine residues or to increase the proline content in order to reduce the free energy of the unfolded state (Margarit et al, 1992; Matthews, 1987a). Other groups concentrated on the importance of additional H-bonds or salt bridges for the stability of a protein (Blaber et al, 1993) or tried to fill cavities in the protein interior to increase the buried hydrophobic surface area and the van der Waals interactions (Karpusas et al, 19898). Furthermore, the stabilization of secondary structure elements, especially a-helices, for example, by improved helix capping, was also investigated (Munoz & Serrano, 1995).

However, there is no fast and promising strategy to identify amino acid replacements which will increase the stability, preferably the thermal stability of a protein. Commonly, the 3D structure of a protein is required to find locations in the molecule where an amino acid replacement possibly will stabilize the protein's folded state. Alternative ways to circumvent this problem are either to search for a homologous protein in a thermo- or hyperthermophile organism or to detect stability-increasing amino acid replacements by a random mutagenesis approach. This latter possibility succeeds in only 10^3 to 10^4 mutations and is restricted to enzymes for which a fast screening procedure is available (Arase et al, 1993; Risse et al, 1992). For all these approaches, success was variable and unpredictable and, if successful, the thermostability enhancements nearly always were rather small.

Here we present an alternative way to improve the thermostability of a protein. Imanaka et al (1986) were among the first to use the comparisons of homologous proteins to enhance the stability of a protein. They used a comparison of proteases from thermophilic with homologous ones of mesophilic organisms to enhance the stability of a mesophilic protease. Serrano et al (1993) used the comparison of the amino acid sequences of two homologous mesophilic RNases to construct a more thermostable Rnase. They mutated individually all of the residues that differ between the two and combined the mutations that increase the stability in a multiple mutant. Pantoliano et al (1989) and, in particular, Steipe et al (1994) suggested that the most frequent amino acid at every position of an alignment of homologous proteins contribute to the largest amount to the stability of a protein. Steipe et al (1994) proved this for a variable domain of an immunoglobulin, whereas Pantoliano et al (1989) looked for positions in the primary sequence of subtilisin in which the sequence of the enzyme chosen to be improved for higher stability was singularly divergent. Their approach resulted in the replacement M50F which increased the T_m of subtilisin by 1.8 °C.

Steipe et al. (1994) proved on a variable domain of immunoglobulin that it is possible to predict a stabilizing mutation with better than 60% success rate just by using a statistical method which determines the most frequent amino acid residue at a certain position of this domain. It was also suggested that this method would provide useful results

not only for stabilization of variable domains of antibodies but also for domains of other proteins. However, it was never mentioned that this method could be extended to the entire protein. Furthermore, nothing is said about the program which was used to calculate the frequency of amino acid residues at a distinct position or whether scoring matrices were used as in the present case.

It is therefore an object of the present invention to provide a process for the preparation of a consensus protein comprising a process to calculate an amino acid residue for nearly all positions of a so-called consensus protein and to synthesize a complete gene from this sequence that could be expressed in a pro- or eukaryotic expression system.

DNA sequences of the present invention can be constructed starting from genomic or cDNA sequences coding for proteins, e.g. phytases known in the art [for sequence information see references mentioned above, e.g.
EP 684 313 or sequence data bases, for example like Genbank (Intelligenetics, California, USA), European Bioinformatics Institute (Hinston Hall, Cambridge, GB), NBRF (Georgetown University, Medical Centre, Washington DC, USA) and Vecbase (University of Wisconsin, Biotechnology Centre, Madison, Wisconsin, USA) or disclosed in the figures by methods of in vitro mutagenesis [see e.g. Sambrook et al., Molecular Cloning, Cold Spring Harbor Laboratory Press, New York]. A widely used strategy for "site directed mutagenesis", as originally outlined by Hurchinson and Edgell [J. Virol. 8, 181 (1971)], involves the annealing of a synthetic oligonucleotide carrying the desired nucleotide substitution to a target region of a single-stranded DNA sequence wherein the mutation should be introduced [for review see Smith, Annu. Rev. Genet. 19, 423 (1985) and for improved methods see references 2-6 in Stanssen et al., Nucl. Acid Res., 17, 4441-4454 (1989)]. Another possibility of mutating a given DNA sequence which is also preferred for the practice of the present invention is the mutagenesis by using the polymerase chain reaction (PCR). DNA as starting material can be isolated by methods known in the art and described e.g. in Sambrook et al. (Molecular Cloning) from the respective strains. For strain information see, e.g. EP 684 313 or any depository authority indicated below. Aspergillus niger [ATCC 9142], Myceliophthora thermophila [ATCC 48102], Talaromyces thermophilus [ATCC 20186] and Aspergillus fumigatus [ATCC 34625] have been redeposited according to the conditions of the Budapest Treaty at the American Type Culture Cell Collection under the following accession numbers: ATCC 74337, ATCC 74340, ATCC 74338 and ATCC 74339, respectively. It is however, understood that DNA encoding a consensus protein in accordance with the present invention can also be prepared in a synthetic manner as described, e.g. in EP 747 483 or the examples by methods known in the art.

The process of the present invention can preferably be used in order to improve the thermostability of the enzyme phytase. After having constructed different consensus phytase sequences it was possible to decide whether single amino acid replacements had a positive or a negative effect on the protein stability. It is therefore another subject of the present invention to improve the thermostability of a phytase.

In this embodiment single amino acids are changed in the sequence of the phytase by the introduction of at least one mutation selected from the group consisting of

E58A	F54Y
D69K	I73V
D197N	K94A
T214L	R101A
E222T	N153K
E267D	V158I
R291I	A203G
R329H	S205G
S364T	V217A
A379K	A227V
G404A	V234L
	P238A
	Q277E
	A287H
	A292Q
	V366I
	A396S
	E415Q
	G437A
	E451R

In the above-given mutations the number represents the position in the consensus phytase-1-sequence as given in Figure 2 and the letter before the number represents the amino acid in the phytase which is replaced by the respective amino acid behind the number. The numbers given correspond to the consensus phytase sequence or relate to a corresponding residue calculated by an alignment as shown in Figure 1 when 26 amino acids (signal sequence) are added to the sequences shown in Fig. 1. Those mutations can be introduced into consensus sequences or into sequences of specific enzymes which have been improved by a process of the present invention. The above-mentioned amino acid replacements have a positive effect on the protein stability.

Once complete DNA sequences of the present invention have been obtained they can be integrated into vectors by methods known in the art and described e.g. in Sambrook et al. (s.a.) to overexpress the encoded polypeptide in appropriate host systems. However, a man skilled in the art knows that also the DNA sequences themselves can be used to transform the suitable host systems of the invention to get overexpression of the encoded polypeptide. Appropriate host systems are for example fungi, like Aspergilli, e.g. Aspergillus niger [ATCC 9142] or Aspergillus ficuum [NRRL 3135] or like Trichoderma, e.g. Trichoderma reesei or yeasts, like Saccharomyces, e.g. Saccharomyces cerevisiae or Pichia, like Pichia pastoris, or Hansenula polymorpha, e.g. H. polymorpha (DSM5215) or plants, as described, e.g. by Pen et al., Bio/Technology 11, 811-814 (1994). A man skilled in the art knows that such microorganisms are available from depository authorities, e.g. the American Type Culture Collection (ATCC), the Centraalbureau voor Schimmelcultures (CBS) or the Deutsche Sammlung für Mikroorganismen und Zellkulturen GmbH (DSM) or any other depository authority as listed in the Journal "Industrial Property" [(1991) 1, pages 29-40]. Bacteria which can be used are e.g. E. coli, Bacilli as, e.g. Bacillus subtilis or Streptomyces, e.g. Streptomyces lividans (see e.g. Anné and Mallaert in FEMS Microbiol. Letters 114, 121 (1993). E. coli, which could be used are E. coli K12 strains e.g. M15 [described as DZ 291 by Villarejo et al. in J. Bacteriol. 120, 466-474 (1974)], HB 101 [ATCC No. 33694] or E. coli SG13009 [Gottesman et al., J. Bacteriol. 148, 265-273 (1981)].

Vectors which can be used for expression in fungi are known in the art and described e.g. in EP 420 358, or by Cullen et al. [Bio/Technology 5, 369-376 (1987)] or Ward in Molecular Industrial Mycology, Systems and Applications for Filamentous Fungi, Marcel Dekker, New York (1991), Upshall et al. [Bio/Technology 5, 1301-1304 (1987)] Gwynne et al. [Bio/Technology 5, 71-79 (1987)], Punt et al. [J. Biotechnol. 17, 19-34 (1991)] and for yeast by Sreekrishna et al. [J. Basic Microbiol. 28, 265-278 (1988), Biochemistry 28,

4117-4125 (1989)], Hitzemann et al. [Nature 293, 717-722 (1981)] or in EP 183 070, EP 183 071, EP 248 227, EP 263 311. Suitable vectors which can be used for expression in E. coli are mentioned, e.g. by Sambrook et al. [s.a.] or by Fiers et al. in Procd. 8th Int. Biotechnology Symposium" [Soc. Franc. de Microbiol., Paris (Durand et al., eds.), pp. 680-697 (1988)] or by Bujard et al. in Methods in Enzymology, eds. Wu and Grossmann, Academic Press, Inc. Vol. 155, 416-433 (1987) and Stüber et al. in Immunological Methods, eds. Lefkovits and Pernis, Academic Press, Inc., Vol. IV, 121-152 (1990). Vectors which could be used for expression in Bacilli are known in the art and described, e.g. in EP 405 370, Procd. Natl. Acad. Sci. USA 81, 439 (1984) by Yansura and Henner, Meth. Enzymol. 185, 199-228 (1990) or EP 207 459. Vectors which can be used for the expression in H. Polymorpha are known in the art and described, e.g. in Gellissen et al., Biotechnology 9, 291-295 (1991).

Either such vectors already carry regulatory elements, e.g. promotors, or the DNA sequences of the present invention can be engineered to contain such elements. Suitable promotor elements which can be used are known in the art and are, e.g. for Trichoderma reesei the cbh1- [Haarki et al., Biotechnology 7, 596-600 (1989)] or the pki1-promotor [Schindler et al., Gene 130, 271-275 (1993)], for Aspergillus oryzae the amy-promotor [Christensen et al., Abstr. 19th Lunteren Lectures on Molecular Genetics F23 (1987), Christensen et al., Biotechnology 6, 1419-1422 (1988), Tada et al., Mol. Gen. Genet. 229, 301 (1991)], for Aspergillus niger the glaA- [Cullen et al., Bio/Technology 5, 369-376 (1987), Gwynne et al., Bio/Technology 5, 713-719 (1987), Ward in Molecular Industrial Mycology, Systems and Applications for Filamentous Fungi, Marcel Dekker, New York, 83-106 (1991)], alcA- [Gwynne et al., Bio/Technology 5, 718-719 (1987)], suc1- [Boddy et al., Curr. Genet. 24, 60-66 (1993)], aphA- [MacRae et al., Gene 71, 339-348 (1988), MacRae et al., Gene 132, 193-198 (1993)], tpiA- [McKnight et al., Cell 46, 143-147 (1986), Upshall et al., Bio/Technology 5, 1301-1304 (1987)], gpdA- [Punt et al., Gene 69, 49-57 (1988), Punt et al., J. Biotechnol. 17, 19-37 (1991)] and the pkiA-promotor [de Graaff et al., Curr. Genet. 22, 21-27 (1992)]. Suitable promotor elements which could be used for expression in yeast are known in the art and are, e.g. the pho5-promotor [Vogel et al., Mol. Cell. Biol., 2050-2057 (1989); Rudolf and Hinnen, Proc. Natl. Acad. Sci. 84, 1340-1344 (1987)] or the gap-promotor for expression in Saccharomyces cerevisiae and for Pichia pastoris, e.g. the aox1-promotor [Koutz et al., Yeast 5, 167-177 (1989); Sreekrishna et al., J. Basic Microbiol. 28, 265-278 (1988)], or the FMD promoter [Hollenberg et al., EPA No. 0299108] or MOX-promotor [Ledeboer et al., Nucleic Acids Res. 13, 3063-3082 (1985)] for H. polymorpha.

Accordingly vectors comprising DNA sequences of the present invention, preferably for the expression of said DNA sequences in bacteria or a fungal or a yeast host and such transformed bacteria or fungal or yeast hosts are also an object of the present invention.

It is also an object of the present invention to provide a system which allows for high expression of proteins, preferably phytases like the consensus phytase of the present invention in Hansenula characterized therein that the codons of the encoding DNA sequence of such a protein have been selected on the basis of a codon frequency table of the organism used for expression, e.g. yeast as in the present case (see e.g. in Example 3) and optionally the codons for the signal sequence have been selected in a manner as described for the specific case in Example 3. That means that a codon frequency table is prepared on the basis of the codons used in the DNA sequences which encode the amino acid sequences of the defined protein family. Then the codons for the design of the DNA sequence of the signal sequence are selected from a codon frequency table of the host cell used for expression whereby always codons of comparable frequency in both tables are used.

Once such DNA sequences have been expressed in an appropriate host cell in a suitable medium the encoded protein can be isolated either from the medium in the case the protein is secreted into the medium or from the host organism in case such protein is present intracellularly by methods known in the art of protein purification or described in case of a phytase, e.g. in EP 420 358. Accordingly a process for the preparation of a polypeptide of the present invention characterized in that transformed bacteria or a host cell as described above is cultured under suitable culture conditions and the polypeptide is recovered therefrom and a polypeptide when produced by such a process or a polypeptide encoded by a DNA sequence of the present invention are also an object of the present invention.

Once obtained the polypeptides of the present invention can be characterized regarding their properties which make them useful in agriculture any assay known in the art and described e.g. by Simons et al. [Br. J. Nutr. 64, 525-540 (1990)], Schöner et al. [J. Anim. Physiol. a. Anim. Nutr. 66, 248-255 (1991)], Vogt [Arch. Geflügelk. 56, 93-98 (1992)], Jongbloed et al. [J. Anim. Sci., 70, 1159-1168 (1992)], Perney et al. [Poultry Sci. 72, 2106-2114 (1993)], Farrell et al., [J. Anim. Physiol. a. Anim. Nutr. 69, 278-283 (1993), Broz et al., [Br. Poultry Sci. 35, 273-280 (1994)] and Düngelhoef et al. [Animal Feed Sci. Technol. 49, 1-10 (1994)] can be used.

In general the polypeptides of the present invention can be used without being limited to a specific field of application, e.g. in case of phytases for the conversion of inositol polyphosphates, like phytate to inositol and inorganic phosphate.

Furthermore the polypeptides of the present invention can be used in a process for the preparation of a pharmaceutical composition or compound food or feeds wherein the components of such a composition are mixed with one or more polypeptides of the present invention. Accordingly compound food or feeds or pharmaceutical compositions comprising one or more polypeptides of the present invention are also an object of the present invention. A man skilled in the art is familiar with their process of preparation. Such pharmaceutical compositions or compound foods or feeds can further comprise additives or components generally used for such purpose and known in the state of the art.

It is furthermore an object of the present invention to provide a process for the reduction of levels of phytate in animal manure characterized in that an animal is fed such a feed composition in an amount effective in converting phytate contained in the feedstuff to inositol and inorganic phosphate.

Before describing the present invention in more detail a short explanation of the Figures enclosed is given below.

Figure 1: Design of the consensus phytase sequence. The letters represent the amino acid residues in the one-letter code. The following sequences were used for the alignment: *phyA* from *Aspergillus terreus* 9A-1 (Mitchell et al, 1997; from amino acid (aa) 27), *phyA* from *A. terreus* cbs116.46; (van Loon et al., 1998; from aa 27), *phyA* from *Aspergillus niger* var. *awamori* (Piddington et al, 1993; from aa 27), *phyA* from *A. niger* T213; from aa 27), *phyA* from *A. niger* strain NRRL3135 (van Hartingsveldt et al, 1993; from aa 27), *phyA* from *Aspergillus fumigatus* ATCC 13073 (Pasamontes et al, 1993; from aa 25), *phyA* from *A. fumigatus* ATCC 32722 (van Loon et al, 1998; from aa 27), *phyA* from *A. fumigatus* ATCC 58128 (van Loon *et al.*, 1998; from aa 27), *phyA* from *A. fumigatus* ATCC 26906 (van Loon et al, 1998; from aa 27), *phyA* from *A. fumigatus* ATCC 32239 (van Loon et al, 1998; from aa 30), *phy*A from *Emericella nidulans* (Pasamontes et al, 1997a; from aa 25), *phy*A from *Talaromyces thermophilus* (Pasamontes et al, 1997a; from aa 24), and *phy*A from *Myceliophthora thermophila* (Mitchell et al, 1997; from aa 19). The alignment was calculated using the program PILEUP. The location of the gaps was refined by hand. Capitalized amino acid residues in the alignment at a given position belong to the amino acid coalition that establish the consensus residue. In bold, beneath the calculated consensus sequence, the amino acid sequence of the finally constructed consensus phytase (Fcp) is shown. The gaps in the calculated consensus sequence were filled by hand according to principals stated in Example 1.

Figure 2: DNA sequence of the consensus phytase-1 gene (*fcp*) and of the primers used for the gene construction. The calculated amino acid sequence (Figure 1) was converted into a DNA sequence using the program BACKTRANSLATE (Devereux *et al.*, 1984) and the codon frequency table of highly expressed yeast genes (GCG program package, 9.0). The signal peptide of the phytase from *A. terreus* cbs.116.46 was fused to the *N*-terminus. The bold bases represent the sequences of the oligonucleotides used to generate the gene. The names of the respective oligonucleotides are alternately noted above or below the sequence. The underlined bases represent the start and stop codon of the gene. The bases written in italics show the two introduced *Eco* RI sites.

Figure 3: Alignment and consensus sequence of five *Basidiomycetes* phytases. The letters represent the amino acid residues in the one-letter code. The amino acid sequences of the phytases from *Paxillus involutus*, phyA1 (aa 21) and phyA2 (aa 21, WO 98/28409), *Trametes pubescens* (aa 24, WO 98/28409), *Agrocybe pediades* (aa 19, WO 98/28409), and *Peniophora lycii* (aa 21, WO 98/28409) starting with the amino acid residues mentioned in parentheses, were used for the alignment and the calculation of the corresponding consensus sequence called "Basidio" (Example 2). The alignment was performed by the program PILEPUP. The location of the gaps was refined by hand. The consensus sequence was calculated by the program PRETTY. While a vote weight of 0.5 was assigned to the two *P. involutus* phytases, all other genes were used with a vote weight of 1.0 for the consensus sequence calculation. At positions, where the program was not able to determine a consensus residues, the Basidio sequence contains a dash. Capitalized amino acid residues in the alignment at a given position belong to the amino acid coalition that establish the consensus residue.

Figure 4: Design of consensus phytase-10 amino acid sequence. Adding the phytase sequence of *Thermomyces lanuginosa* (Berka *et al.*, 1998) and the consensus sequence of the phytases from five *Basidiomycetes* to the alignment of Figure 1, an improved consensus sequence was calculated by the program PRETTY. Additionally, the amino acid sequence of *A. niger* T213 was omitted, therefore, using a vote weight of 0.5 for the remaining *A. niger* phytase sequences. For further information see Example 2.

Figure 5: DNA and amino acid sequence of consensus phytase-10. The amino acid sequence is written above the corresponding DNA sequence using the one-letter code. The sequence of the oligonucleotides which were used to assemble the gene are in bold letters. The label of oligonucleotides and the amino acids, which were changed compared to those for consensus phytase -1, are underlined and their corresponding triplets are highlighted in

small cases. The *fcp*10 gene was assembled from the following oligonucleotides: CP-1, CP-2, CP-3.10, CP-4.10, CP-5.10, CP-6, CP-7.10, CP-8.10, CP-9.10, CP-10.10, CP-11.10, CP-12.10, CP-13.10, CP-14.10, CP-15.10, CP-16.10, CP-17.10, CP18.10, CP-19.10, CP-20.10, CP-21.10, CP-22.10. The newly synthesized oligonucleotides are additionally marked by number 10. The phytase contains the following 32 exchanges: Y54F, **E58A**, D69K, D70G, A94K, N134Q, I158V, S187A, Q188N, **D197N**, S204A, T214L, D220E, L234V, A238P, D246H, T251N, Y259N, **E267D**, E277Q, A283D, **R291I**, A320V, **R329H**, **S364T**, I366V, **A379K**, S396A, **G404A**, Q415E, A437G, A463E. The mutations accentuated in bold letters revealed a stabilizing effect on consensus phytase-1 as tested as single mutation in consensus phytase-1.

Figure 6: Alignment for the design of consensus phytase-11. In contrast to the design of consensus phytase-10, for the design of the amino acid sequence of consensus phytase-11, all *Basidiomycetes* phytases were used as independent sequences using an assigned vote weight of 0.2 for each *Basidiomycetes* sequence. Additionally, the amino acid sequence of *A. niger* T213 was used in that alignment, again.

Figure 7: DNA and amino acid sequence of consensus phytase-1-thermo[8]-Q50T-K91A. The amino acid sequence is written above the corresponding DNA sequence using the one-letter code. The replaced amino acid residues are underlined. The stop codon of the gene is marked by a star (*).

Figure 8: DNA and amino acid sequence of consensus phytase-10-thermo[3]-Q50T-K91A. The amino acid sequence is written above the corresponding DNA sequence using the one-letter code. The replaced amino acid residues are underlined. The stop codon of the gene is marked by a star (*).

Figure 9: DNA and amino acid sequence of *A. fumigatus* ATCC 13073 phytase a-mutant. The amino acid sequence is written above the corresponding DNA sequence using the one-letter code. The replaced amino acid residues are underlined. The stop codon of the gene is marked by a star (*).

Figure 10: DNA and amino acid sequence of consensus phytase-7. The amino acids are written above the corresponding DNA sequence using the one-letter code. The sequence of the oligonucleotides used to assemble the gene are in bold letters. Oligonucleotides and amino acids that were exchanged are underlined and their corresponding triplets are highlighted in small cases. The *fcp*7 gene was assembled from the following

oligonucleotides: CP-1, CP-2, CP-3, CP-4.7, CP-5.7, CP-6, CP-7, CP-8.7, CP-9, CP-10.7, CP-11.7, CP-12.7, CP-13.7, CP-14.7, CP-15.7, CP-16, CP-17.7, CP-18.7, CP-19.7, CP-20, CP-21, CP-22. The newly synthesized oligonucleotides are additionally marked by number 7. The phytase contains the following 24 exchanges in comparison to the original consensus phytase: S89D, S92G, A94K, D164S, P201S, G203A, G205S, H212P, G224A, D226T, E255T, D256E, V258T, P265S, Q292H, G300K, Y305H, A314T, S364G, M365I, A397S, S398A, G404A, and A405S.

Figure 11: Differential scanning calorimetry (DSC) of consensus phytase-1 and consensus phytase-10. The protein samples were concentrated to ca. 50-60 mg/ml and extensively dialyzed against 10 mM sodium acetate, pH 5.0. A constant heating rate of 10 °C/min was applied up to 95 °C. DSC of consensus phytase-10 (upper graph) yielded a melting temperature of 85.4 °C, which is 7.3 °C higher than the melting point of consensus phytase-1 (78.1 °C, lower graph).

Figure 12: Differential scanning calorimetry (DSC) of consensus phytase-10-thermo-Q50T and consensus phytase-10-thermo-Q50T-K91A. The protein samples were concentrated to ca. 50-60 mg/ml and extensively dialyzed against 10 mM sodium acetate, pH 5.0. A constant heating rate of 10 °C/min was applied up to 95 °C. DSC of consensus phytase-10-thermo-Q50T (upper graph) yielded a melting temperature of 88.6 °C, while the melting point of consensus phytase-10-thermo-Q50T-K91A was found at 89.3 °C.

Figure 13: Comparison of the temperature optimum between consensus phytase-1, consensus phytase-10 and consensus phytase-10-thermo-Q50T. For the determination of the temperature optimum, the phytase standard assay was performed at a series of temperatures between 37 and 86 °C. The diluted supernatant of transformed *S. cerevisiae* strains was used for the determination. The other components of the supernatant showed no influence on the determination of the temperature optimum: ∧, consensus phytase-1; ◇, consensus phytase-10; ■, consensus phytase 10-thermo-Q50T.

Figure 14: pH-dependent activity profile and substrate specificity of consensus phytase-10 and its variants thermo-Q50T and thermo-Q50T-K91A. The phytase activity was determined using the standard assay in appropriate buffers (see Example 9) at different pH-values. Graph a) shows the pH-dependent activity profile of consensus phytase-10 (□), consensus phytase-10-thermo-Q50T (•), and consensus phytase-10-thermo-Q50T-K91A (∧). Graph b) shows the corresponding substrate specificity tested by replacement of phytate by the indicated compounds in the standard assay; open bars, consensus phytase-10

(grey bars, consensus phytase-10-thermo-Q50T; dark bars, consensus phytase-10-thermo-Q50T-K91A). The numbers correspond to the following compounds: 1, phytate; 2, *p*-nitrophenyl phosphate; 3, phenyl phosphate; 4, fructose-1,6-bisphosphate; 5, fructose-6-phosphate; 6, glucose-6-phosphate; 7, ribose-5-phosphate; 8, DL-glycerol-3-phosphate; 9, glycerol-2-phosphate; 10, 3-phosphoglycerate; 11, phosphoenolpyruvate; 12, AMP; 13, ADP; 14, ATP.

Figure 15: pH-dependent activity profile and substrate specificity of consensus phytase-1-thermo[8]-Q50T and of consensus phytase-1-thermo[8]-Q50T-K91A. The phytase activity was determined using the standard assay in appropriate buffers (see Example 9) at different pH-values. Graph a) shows the pH-dependent activity profile of the Q50T- (■) and the Q50T-K91A-variant (•). Graph b) shows the corresponding substrate specificities tested by replacement of phytate by the indicated compounds in the standard assay (open bars, consensus phytase-1-thermo[8]-Q50T; filled bars, consensus phytase-1-thermo[8]-Q50T-K91A.). The substrates are listed in the legend of Figure 14.

Figure 16: Differential scanning calorimetry (DSC) of consensus phytase-1-thermo[8]-Q50T and consensus phytase-1-thermo[8]-Q50T-K91A. The protein samples were concentrated to ca. 50-60 mg/ml and extensively dialyzed against 10 mM sodium acetate, pH 5.0. A constant heating rate of 10 °C/min was applied up to 95 °C. DSC of consensus phytase-1-thermo[8]-Q50T (upper graph) showed a melting temperature of 84.7 °C, while the melting point of consensus phytase-1-thermo[8]-Q50T-K91A was found at 85.7 °C.

Figure 17: Comparison of the temperature optimum between consensus phytase-1, consensus phytase-1-thermo[3] and consensus phytase-1-thermo[8]. For the determination of the temperature optimum, the phytase standard assay was performed at a series of temperatures between 37 and 86 °C. Purified protein from the supernatant of transformed *S. cerevisiae* strains was used for the determination. O, consensus phytase-1; □, consensus phytase-1-thermo[3]; ▲ , consensus phytase 1-thermo[8].

Figure 18: Comparison of the pH-dependent activity profile and substrate specificity of consensus phytase-1, consensus phytase-7, and of the phytase from *A. niger* NRRL 3135. The phytase activity was determined using the standard assay in appropriate buffers (see Example 9) at different pH-values. Graph a) shows the pH-dependent activity profile of consensus phytase-1 (■), the phytase from *A. niger* NRRL 3135 (O), and of consensus phytase-7 (▲). Graph b) shows the corresponding substrate specificity tested by replacement of phytate by the indicated compounds in the standard assay (black bars, *A.*

niger NRRL 3135 phytase; grey bars, consensus phytase-1, dashed bars, consensus phytase-7). The substrates are listed in the legend of Figure 14.

Figure 19: Differential scanning calorimetry (DSC) of the phytase from A. fumigatus ATCC 13073 and of its stabilized α-mutant, which contains the following amino acid exchanges F55Y, V100I, F114Y, A243L, S265P, N294D.

The protein samples were concentrated to ca. 50-60 mg/ml and extensively dialyzed against 10 mM sodium acetate, pH 5.0. A constant heating rate of 10 °C/min was applied up to 95 °C. DSC of consensus *A. fumigatus* 13073 phytase (upper graph) revealed a melting temperature of 62.5 °C, while the melting point of the α-mutant was found at 67.0 °C

Figure 20: Comparison of the temperature optimum of *A. fumigatus* 13073 wild-type, its *A. fumigatus* α-mutant, and a further stabilized α-mutant (E59A-S126N-R329H-S364T-G404A). For the determination of the temperature optimum, the phytase standard assay was performed at a series of temperatures between 37 and 75 °C. The diluted supernatant of transformed *S. cerevisiae* strains was used for the determination. The other components of the supernatant showed no influence on the determination of the temperature optimum. ○, *A. fumigatus* ATCC 13073 phytase; ▲, *A. fumigatus* ATCC 13073 α-mutant; □, *A. fumigatus* ATCC 13073 alpha-mutant-(E59A-S126N-R329H-S364T-G404A)-Q27T; ■, *A. fumigatus* ATCC 13073 α-mutant-(E59A-S126N-R329H-S364T-G404A)-Q27T-K68A. Q27T and K68A corresponds to consensus phytase-1 Q50T and K91A, respectively.

Figure 21: Amino acid sequence of consensus phytase 12 (consphy12) which contains a number of active site residues transferred from the "basidio" consensus sequence to consensus phytase-10-thermo-Q50T-K91A.

Example 1:
Design of the amino acid sequence of consensus phytase-1

Alignment of the amino acid sequences

The alignment was calculated using the program PILEUP from the Sequence Analysis Package Release 9.0 (Devereux *et al.*, 1984) with the standard parameter (gap creation penalty 12, gap extension penalty 4). The location of the gaps was refined using a text editor. Table 1 shows the sequences (see Figure 1) without the signal sequence that were used for the performance of the alignment starting with the amino acid (aa) as mentioned in Table 1.

Table 1: Origin and vote weight of the phytase amino acid sequences used for the design of consensus phytase-1

- *phyA* from *Aspergillus terreus* 9A-1, aa 27, vote weight 0.5 (Mitchell *et al.*, 1997)
- *phyA* from *Aspergillus terreus* cbs116.46, aa 27, vote weight 0.5 (van Loon *et al.*, 1998)
- *phyA* from *Aspergillus niger* var. *awamori*, aa 27, vote weight 0.33 (Piddington *et al.*, 1993)
- *phyA* from *Aspergillus niger* T213, aa 27, vote weight 0.33
- *phyA* from *Aspergillus niger* strain NRRL3135, aa 27, vote weight 0.33 (van Hartingsveldt *et al.*, 1993)
- *phyA* from *Aspergillus fumigatus* ATCC 13073, aa 26, vote weight 0.2 (Pasamontes *et al.*, 1997)
- *phyA* from *Aspergillus fumigatus* ATCC 32722, aa 26, vote weight 0.2 (van Loon *et al.*, 1998)
- *phyA* from *Aspergillus fumigatus* ATCC 58128, aa 26, vote weight 0.2 (van Loon *et al.*, 1998)
- *phyA* from *Aspergillus fumigatus* ATCC 26906, aa 26, vote weight 0.2 (van Loon *et al.*, 1998)
- *phyA* from *Aspergillus fumigatus* ATCC 32239, aa 30, vote weight 0.2 (van Loon *et al.*, 1998)
- *phyA* from *Emericella nidulans* , aa 25, vote weight 1.0 (Roche Nr. R1288, Pasamontes *et al.*, 1997a)
- *phyA* from *Talaromyces thermophilus* ATCC 20186, aa 24, vote weight 1.0 (Pasamontes *et al.*, 1997a)
- *phyA* from *Myceliophthora thermophila*, aa 19, vote weight 1.0 (Mitchell *et al.*, 1997)

Calculation of the amino acid sequence of consensus phytase-1

Using the refined alignment as input, the consensus sequence was calculated by the program PRETTY from the Sequence Analysis Package Release 9.0 (Devereux *et al.*, 1984). PRETTY prints sequences with their columns aligned and can display a consensus sequence for an alignment. A vote weight that pays regard to the similarity between the amino acid sequences of the phytases aligned was assigned to all sequences. The vote weight was set such as the combined impact of all phytases from one sequence subgroup (same species, but from different strains), e. g. the amino acid sequences of all phytases from *A. fumigatus*, on the election was set one, that means that each sequence contributes with a value of 1 divided by the number of strain sequences (see Table 1). By this means, it was possible to prevent that very similar amino acid sequences, e. g. of the phytases from different *A. fumigatus* strains, dominate the calculated consensus sequence.

The program PRETTY was started with the following parameters: The plurality defining the number of votes below which there is no consensus was set on 2.0. The threshold, which determines the scoring matrix value below which an amino acid residue may not vote for a coalition of residues, was set on 2. PRETTY used the PrettyPep.Cmp consensus scoring matrix for peptides.

Ten positions of the alignment (position 46, 66, 82, 138, 162, 236, 276, 279, 280, 308; Figure 1), for which the program was not able to determine a consensus residue, were filled by hand according to the following rules: if a most frequent residue existed, this residue was chosen (138, 236, 280); if a prevalent group of similar or phylogenetically equivalent residues occurred, the most frequent or, if not available, one residues of this group was selected (46, 66, 82, 162, 276, 308). If there was either a prevalent residue nor a prevalent group, one of the occurring residues was chosen according to common assumption on their influence on the protein stability (279). Eight other positions (132, 170, 204, 211, 275, 317, 384, 447; Figure 1) were not filled with the amino acid residue selected by the program but normally with amino acids that occur with the same frequency as the residues that were chosen by the program. In most cases, the slight underrating of the three *A. niger* sequences (sum of the vote weights: 0.99) was eliminated by this corrections.

Conversion of the consensus phytase-1 amino acid sequence to a DNA sequence

The first 26 amino acid residues of *A. terreus* cbs116.46 phytase were used as signal peptide and, therefore, fused to the N-terminus of all consensus phytases. For this stretch,

we used a special method to calculate the corresponding DNA sequence. Purvis et al (1987) proposed that the incorporation of rare codons in a gene has an influence on the folding efficiency of the protein. Therefore, at least the distribution of rare codons in the signal sequence of *A. terreus* cbs116.46, which was used for the consensus phytase and which is very important for secretion of the protein, but converted into the *S. cerevisiae* codon usage, was transferred into the new signal sequence generated for expression in *S. cerevisiae*. For the remaining parts of the protein, we used the codon frequency table of highly expressed *S. cerevisiae* genes, obtained from the GCG program package, to translate the calculated amino acid sequence into a DNA sequence.

The resulting sequence of the *fcp* gene is shown in Figure 2.

Construction and cloning of the consensus phytase-1 gene

The calculated DNA sequence of consensus phytase-1 (*fcp*) was divided into oligonucleotides of 85 bp, alternately using the sequence of the sense and the anti-sense strand. Every oligonucleotide overlaps 20 bp with its previous and its following oligonucleotide of the opposite strand. The location of all primers, purchased by Microsynth, Balgach (Switzerland) and obtained in a PAGE-purified form, is indicated in Figure 2.

PCR-Reactions

In three PCR reactions, the synthesized oligonucleotides were composed to the entire gene. For the PCR, the High Fidelity Kit from Boehringer Mannheim (Boehringer Mannheim, Mannheim, Germany) and the thermo cycler The Protokol™ from AMS Biotechnology (Europe) Ltd. (Lugano, Switzerland) was used.

Oligonucleotide CP-1 to CP-10 (Mix 1, Figure 2) were mixed to a concentration of 0.2 pMol/µl of each oligonucleotide. A second oligonucleotide mixture (Mix 2) was prepared with CP-9 to CP-22 (0.2 pMol/µl of each oligonucleotide). Additionally, four short primers were used in the PCR reactions:

CP-a: *Eco* RI

5'-TATAT*GAATTC*ATGGGCGTGTTCGTC-3'

CP-b:

5'-TGAAAAGTTCATTGAAGGTTTC-3'

CP-c:

5'-TCTTCGAAAGCAGTACAAGTAC-3'

CP-e: *Eco* RI

5'-TATAT*GAATTC*TTAAGCGAAAC-3'

PCR reaction *a*:
10 µl Mix 1 (2.0 pmol of each oligonucleotide)
2 µl nucleotides (10 mM each nucleotide)
2 µl primer CP-a (10 pmol/µl)
2 µl primer CP-c (10 pmol/µl)
10,0 µl PCR buffer
0.75 µl polymerase mixture
73.25 µl H_2O

PCR reaction *b*:
10 µl Mix 2 (2.0 pmol of each oligonucleotide)
2 µl nucleotides (10 mM each nucleotide)
2 µl primer CP-b (10 pmol/µl)
2 µl primer CP-e (10 pmol/µl)
10,0 µl PCR buffer
0.75 µl polymerase mixture (2.6 U)
73.25 µl H_2O

Reaction conditions for PCR reaction *a* and *b*:

step 1	2 min - 45°C
step 2	30 sec - 72°C
step 3	30 sec - 94°C
step 4	30 sec - 52°C
step 5	1 min - 72°C

Step 3 to 5 were repeated 40-times.

The PCR products (670 and 905 bp) were purified by an agarose gel electrophoresis (0.9% agarose) and a following gel extraction (QIAEX II Gel Extraction Kit, Qiagen, Hilden, Germany). The purified DNA fragments were used for the PCR reaction *c*.

PCR reaction *c*:
6 µl PCR product of reaction a (≈50 ng)
6 µl PCR product of reaction b (≈50 ng)
2 µl primer CP-a (10 pmol/µl)
2 µl primer CP-e (10 pmol/µl)
10,0 µl PCR buffer
0.75 µl polymerase mixture (2.6 U)
73.25 µl H_2O

Reaction conditions for PCR reaction *c*:

step 1 2 min - 94°C
step 2 30 sec - 94°C
step 3 30 sec - 55°C
step 4 1 min - 72°C

Step 2 to 4 were repeated 31-times.

The resulting PCR product (1.4 kb) was purified as mentioned above, digested with *Eco* RI, and ligated in an *Eco* RI-digested and dephosphorylated pBsk(-)-vector (Stratagene, La Jolla, CA, USA). 1 µl of the ligation mixture was used to transform *E. coli* XL-1 competent cells (Stratagene, La Jolla, CA, USA). All standard procedures were carried out as described by Sambrook *et al.* (1987). The DNA sequence of the constructed consensus phytase gene (*fcp*, Figure 2) was controlled by sequencing as known in the art.

Example 2
Design of an improved consensus phytase (consensus phytase-10) amino acid sequence

The alignments used for the design of consensus phytase-10 were calculated using the program PILEUP from the Sequence Analysis Package Release 9.0 (Devereux *et al.*, 1984) with the standard parameter (gap creation penalty 12, gap extension penalty 4). The location of the gaps was refined using a text editor.

The following sequences were used for the alignment of the *Basiodiomycetes* phytases starting with the amino acid (aa) mentioned in Table 2:

Table 2: Origin and vote weight of five *Basidiomycetes* phytases used for the calculation of the corresponding amino acid consensus sequence (basidio)

- *phyA1* from *Paxillus involutus* NN005693, aa 21, vote weight 0.5 (WO 98/28409)
- *phyA2* from *Paxillus involutus* NN005693, aa 21, vote weight 0.5 (WO 98/28409)
- *phyA* from *Trametes pubescens* NN9343, aa 24, vote weight 1.0 (WO 98/28409)
- *phyA* from *Agrocybe pediades* NN009289, aa 19, vote weight 1.0 (WO 98/28409)
- *phyA* from *Peniophora lycii* NN006113, aa 21, vote weight 1.0 (WO 98/28409)

The alignment is shown in Figure 3.

In Table 3 the genes, which were used for the performance of the final alignment, are arranged. The first amino acid (aa) of the sequence which is used in the alignment is mentioned behind the organism designation.

Table 3: Origin and vote weight of the phytase sequences used for the design of consensus phytase 10

- *phyA* from *Aspergillus terreus* 9A-1, aa 27, vote weight 0.5 (Mitchell *et al.*, 1997)
- *phyA* from *Aspergillus terreus* cbs116.46, aa 27, vote weight 0.5 (van Loon *et al.*, 1998)
- *phyA* from *Aspergillus niger* var. *awamori*, aa 27, vote weight 0.5 (Piddington *et al.*, 1993)
- *phyA* from *Aspergillus niger* strain NRRL3135, aa 27, vote weight 0.5 (van Hartingsveldt *et al.*, 1993)
- *phyA* from *Aspergillus fumigatus* ATCC 13073, aa 26, vote weight 0.2 (Pasamontes *et al.*, 1997)
- *phyA* from *Aspergillus fumigatus* ATCC 32722, aa 26, vote weight 0.2 (van Loon *et al.*, 1998)
- *phyA* from *Aspergillus fumigatus* ATCC 58128, aa 26, vote weight 0.2 (van Loon *et al.*, 1998)
- *phyA* from *Aspergillus fumigatus* ATCC 26906, aa 26, vote weight 0.2 (van Loon *et al.*, 1998)
- *phyA* from *Aspergillus fumigatus* ATCC 32239, aa 30, vote weight 0.2 (van Loon *et al.*, 1998)
- *phyA* from *Emericella nidulans* , aa 25, vote weight 1.0 (Roche Nr. R1288, Pasamontes *et al.*, 1997a)
- *phyA* from *Talaromyces thermophilus* ATCC 20186, aa 24, vote weight 1.0 (Pasamontes *et al.*, 1997a)
- *phyA* from *Myceliophthora thermophila*, aa 19, vote weight 1.0 (Mitchell *et al.*, 1997)
- *phyA* from *Thermomyces lanuginosa*, aa 36, vote weight 1.0 (Berka *et al.*, 1998)
- Consensus sequence of five *Basidiomycetes* phytases, vote weight 1.0 (Basidio, Figure 3)

The corresponding alignment is shown in Figure 4.

Calculation of the amino acid sequence of consensus-10

To improve the alignment, we added the original consensus sequence of five phytases from four different *Basidiomycetes*, called Basidio, still containing the undefined sequence positions (see Figure 3), nearly all phytase sequences used for calculation of the original consensus phytase and one new phytase sequence from the *Ascomycete Thermomyces lanuginosa* to a larger alignment. Using the consensus sequence of the basidiomycetal phytase sequences, does not pay regard to the diversity among the five amino acid sequences, but pays regard to the common and different amino acid residues between the phytases from the *Ascomycetes* and the *Basidiomycetes*.

We set plurality on 2.0 and threshold on 3. The used vote weight are listed in Table 3. The alignment and the corresponding consensus sequence is presented in Figure 4. The new consensus phytase sequence has 32 different amino acids in comparison to the original consensus phytase. Positions for which the program PRETTY was not able to calculate a consensus amino acid residue were filled according to rules mentioned in Example 1. None of the residues suggested by the program was replaced.

Furthermore, we included all *Basidiomycetes* phytases as single amino acid sequences but assigning a vote weight of 0.2 in the alignment. The corresponding alignment is shown in Figure 6. The calculated consensus amino acid sequence (consensus phytase-11) has the following differences to the sequence of consensus phytase-10. Letter X means that the program was not able to calculate a consensus amino acid; the amino acid in parenthesis corresponds to the amino acid finally included into the consensus phytase-10.

D35X, X(K)69K, X(E)100E, A101R, Q134N, X(K)153N, X(H)190H, X(A)204S, X(E)220D, E222T, V227A, X(R)271R, H287A, X(D)288D, X(K)379K, X(I)389I, E390X, X(E)415E, X(A)416A, X(R)446L, E463A, whereas the numbering is as in Fig. 5.

We also checked single amino acid replacements suggested by the improved consensus sequences 10 and 11 on their influence on the stability of the original consensus phytase. The approach is described in example 3.

Conversion of consensus phytase-10 amino acid sequence to a DNA sequence

The first 26 amino acid residues of *A. terreus* cbs116.46 phytase were used as signal peptide and, therefore, fused to the *N*-terminus of consensus phytase-10. The used procedure is further described in Example 1.

The resulting sequence of the *fcp*10 gene is shown in Figure 5.

Construction and cloning of the consensus phytase-10 gene (*fcp*10)

The calculated DNA sequence of *fcp*10 was divided into oligonucleotides of 85 bp, alternately using the sequence of the sense and the anti-sense strand. Every oligonucleotide overlaps 20 bp with its previous and its following oligonucleotide of the opposite strand. The location of all primers, purchased by Microsynth, Balgach (Switzerland) and obtained in a PAGE-purified form, is indicated in Figure 5.

PCR-Reactions

In three PCR reactions, the synthesized oligonucleotides were composed to the entire gene. For the PCR, the High Fidelity Kit from Boehringer Mannheim (Boehringer Mannheim, Mannheim, Germany) and the thermo cycler The Protokol™ from AMS Biotechnology (Europe) Ltd. (Lugano, Switzerland) was used. The following oligonucleotides were used in a concentration of 0.2 pMol/ml.

Mix 1.10: CP-1, CP-2, CP-3.10, CP-4.10, CP-5.10, CP-6, CP-7.10, CP-8.10, CP-9.10, CP-10.10

Mix 2.10: CP-9.10, CP-11.10, CP-12.10, CP-13.10, CP-14.10, CP-15.10, CP-16.10, CP-17.10, CP18.10, CP-19.10, CP-20.10, CP-21.10, CP-22.10

The newly synthesized oligonucleotides are marked by number 10. The phytase contains the following 32 exchanges, which are underlined in Figure 5, in comparison to the original consensus phytase: Y54F, E58A, D69K, D70G, A94K, N134Q, I158V, S187A, Q188N, D197N, S204A, T214L, D220E, L234V, A238P, D246H, T251N, Y259N, E267D, E277Q, A283D, R291I, A320V, R329H, S364T, I366V, A379K, S396A, G404A, Q415E, A437G, A463E.

Four short PCR primer were used for the assembling of the oligonucleotides:

CP-a: *Eco* RI
5'-TATAT*GAATTC*ATGGGCGTGTTCGTC-3'

CP-b:
5'-TGAAAAGTTCATTGAAGGTTTC-3'

CP-c.10:
5'-TCTTCGAAAGCAGTACACAAAC-3'

CP-e: *Eco* RI
5'-TATAT*GAATTC*TTAAGCGAAAC-3'

PCR reaction *a*:
10 µl Mix 1.10 (2.0 pmol of each oligonucleotide)
2 µl nucleotides (10 mM each nucleotide)
2 µl primer CP-a (10 pmol/ml)
2 µl primer CP-c.10 (10 pmol/ml)
10,0 µl PCR buffer
0.75 µl polymerase mixture
73.25 µl H_2O

PCR reaction *b*:

10 µl Mix 2.10 (2.0 pmol of each oligonucleotide)
2 µl nucleotides (10 mM each nucleotide)
2 µl primer CP-b (10 pmol/ml)
2 µl primer CP-e (10 pmol/ml)
10,0 µl PCR buffer
0.75 µl polymerase mixture (2.6 U)
73.25 µl H_2O

Reaction conditions for PCR reaction *a* and *b*:

step 1	2 min - 45 °C
step 2	30 sec - 72 °C
step 3	30 sec - 94 °C
step 4	30 sec - 52 °C
step 5	1 min - 72 °C

Step 3 to 5 were repeated 40-times.

The PCR products (670 and 905 bp) were purified by an agarose gel electrophoresis (0.9% agarose) and a following gel extraction (QIAEX II Gel Extraction Kit, Qiagen, Hilden, Germany). The purified DNA fragments were used for the PCR reaction *c*.

PCR reaction *c*:

6 µl PCR product of reaction a ≈50 ng)
6 µl PCR product of reaction b ≈50 ng)
2 µl primer CP-a (10 pmol/ml)
2 µl primer CP-e (10 pmol/ml)
10,0 µl PCR buffer
0.75 µl polymerase mixture (2.6 U)
73.25 µl H_2O

Reaction conditions for PCR reaction *c*:

step 1	2 min - 94 °C
step 2	30 sec - 94 °C
step 3	30 sec - 55 °C
step 4	1 min - 72 °C

Step 2 to 4 were repeated 31-times.

The resulting PCR product (1.4 kb) was purified as mentioned above, digested with *Eco* RI, and ligated in an *Eco* RI-digested and dephosphorylated pBsk(-)-vector (Stratagene, La Jolla, CA, USA). 1 µl of the ligation mixture was used to transform *E. coli* XL-1 competent cells (Stratagene, La Jolla, CA, USA). All standard procedures were carried out as described by Sambrook *et al.* (1987). The DNA sequence of the constructed gene (*fcp10*) was checked by sequencing as known in the art.

Example 3

Increasing the thermostability of consensus phytase-1 by introduction of single mutations suggested by the amino acid sequence of consensus phytase-10 and consensus phytase-11

In order to increase the thermostability of homologous genes, it is also possible to test the stability effect of each differing amino acid residue between the protein of interest and the calculated consensus sequence and to combine all stabilizing mutations into the protein of interest. We used the consensus phytase as protein of interest and tested the effect on the protein stability of 34 amino acid residues, differing to consensus phytase 10 and/or 11 as single mutations.

To construct muteins for expression in *A. niger*, *S. cerevisiae*, or *H. polymorpha*, the corresponding expression plasmid containing the consensus phytase gene was used as template for site-directed mutagenesis (see Example 6-8). Mutations were introduced using the "quick exchange™ site-directed mutagenesis kit" from Stratagene (La Jolla, CA, USA) following the manufacturer's protocol and using the corresponding primers. All mutations made and their corresponding primers are summarized in Table 4. Plasmids harboring the desired mutation were identified by DNA sequence analysis as known in the art.

Table 4: Primers used for site-directed mutagenesis of consensus phytase

(Exchanged bases are highlighted in bold. The introduction of a restriction site is marked above the sequence. When a restriction site is written in parenthesis, the mentioned site was destroyed by introduction of the mutation.)

mutation	Primer set
	Kpn I
Q50T	5'-CACTTGTGG***GGTACC***TACTCTCCATACTTCTC-3'
	5'-GAGAAGTATGGAGAGTA***GGTAC***CCCACAAGTG-3'
Y54F	5'-GGTCAATACTCTCCATTCTTCTCTTTGGAAG-3'
	5'-CTTCCAAAGAGAAGAATGGAGAGTATTGACC-3'
E58A	5'-CATACTTCTCTTTGGCAGACGAATCTGC-3'
	5'-GCAGATTCGTCTGCCAAAGAGAAGTATG-3'

		Aat II
	D69K	5'-CTCCA*GACGTC*CCAAA**G**GACTGTAGAGTTAC-3'
		5'-GTAACTCTACAGTC**CTT**TGG*GACGTC*TGGAG-3'
		Aat II
	D70G	5'-CTCCA*GACGTC*CCAGACG**G**CTGTAGAGTTAC-3'
		5'-GTAACTCTACAG**C**CGTCTGG*GACGTC*TGGAG-3'
	K91A	5'-GATACCCAACTTCTTCT**G**CGTCTAAGGCTTACTCTG-3'
		5'-CAGAGTAAGCCTTAGAC**G**CAGAAGAAGTTGGGTATC-3'
		Sca I
	A94K	5'-CTTCTAAGTCTAAGA*AGTACT*CTGCTTTG-3'
		5'-CAAAGCAG*AGTACT*TCTTAGACTTAGAAG-3'
	A101R	5'-GCTTACTCTGCTTTGATTGAA**CGG**ATTCAAAAGAACGCTAC-3'
		5'-GTAGCGTTCTTTTGAAT**CCG**TTCAATCAAAGCAGAGTAAGC-3'
	N134Q	5'-CCATTCGGTGAA**C**AG**C**AAATGGTTAACTC-3'
		5'-GAGTTAACCATTTG**C**T**G**TTCACCGAATGG-3'
		Nru I
	K153N	5'-GATACAAGGCT*CTCGCGA*GAAA**C**ATTGTTC –3'
		5'-GGAACAAT**G**TTTC*TCGCGA***G**AGCCTTGTATC-3'
		Bss HI
	I158V	5'-GATTGTTCCATTCGT*GCGCGC*TTCTGGTTC-3'
		5'-GAACCAGAA*GCGCGC*ACGAATGGAACAATC-3'
		Bcl I
	D197N	5'-CTCCAGTTATT**AACG***TGATCA*TTCCAGAAGG-3'
		5'-CCTTCTGGAA*TGATCA*C**GTT**AATAACTGGAG-3'
		Apa I
	S187A	5'-GGCTGACCCAGG***GGCCC***AACCACACCAAGC-3'
		5'-GCTTGGTGTGGTT*GGGCCC*CTGGGTCAGCC-3'
		Nco I
	T214L	5'-CACTTTGGA*CCATGG*T**C**TTTGTACTGCTTTCG-3'
		5'-CGAAAGCAGTACAAA**G**A*CCATGG*TCCAAAGTG-3'
		Avr II
	E222T	5'-GCTTTCGAAGACTCTA**C***CCTAGG*TGACGACGTTG-3'
		5'-CAACGTCGTCA*CCTAGG***GG**TAGAGTCTTCGAAAGC-3'

	V227A	5'-GGTGACGACGCTGAAGCTAACTTCAC-3' 5'-GTGAAGTTAGCTTCAGCGTCGTCACC-3'
		Sac II
5	L234V	5'-CTAACTTCA*CCGCGG*TGTTCGCTCCAG-3' 5'-CTGGAGCGAACACCGCGGTGAAGTTAG-3'
	A238P	5'-GCTTTGTTCGCTCCACCTATTAGAGCTAGATTGG-3' 5'-CCAATCTAGCTCTAATAGGTGGAGCGAACAAAGC-3'
10		*Hpa* I
	T251N	5'-GCCAGGT*GTTAAC*TTGACTGACGAAG-3' 5'-TTCGTCAGTCAA*GTTAAC*ACCTGGC-3'
		Aat II
	Y259N	5'-GACGAA*GACGTC*GTTAACTTGATGGAC-3'
15		5'-GTCCATCAAGTTAAC*GACGTC*TTCGTC-3'
		Asp I
	E267D	5'-GTCCATTC*GACACTGTC*GCTAGAACTT C-3' 5'-GAAGTTCTAGC*GACAGTGT*CGAATGGAC-3'
20	E277Q	5'-CTGACGCTACTCAGCTGTCTCCATTC-3' 5'-GAATGGAGACAGCTGAGTAGCGTCAG-3'
	A283D	5'-GTCTCCATTCTGTGATTTGTTCACTCAC-3' 5'-GTGAGTGAACAAATCACAGAATGGAGAC-3'
25		*Ksp* I
	H287A	5'-GCTTTGTTCA*CCGCGG*ACGAATGGAG-3' 5'-CTCCATTCGT*CCGCGG*TGAACAAAGC-3'
		Bam HI
	R291I	5'-CACGACGAAT*GGATCC*AATACGACTAC-3'
30		5'-GTAGTCGTATT*GGATCC*ATTCGTCGTG-3'
		Bsi WI
	Q292A	5'-GACGAATGGAGA*GCGTACG*ACTACTTG-3' 5'-CAAGTAGT*CGTACGC*TCTCCATTCGTC-3'
		Hpa I
35	A320V	5'-GGTGTTGGTTTC*GTTAAC*GAATTGATTGC-3' 5'-GCAATCAATTC*GTTAAC*GAAACCAACACC-3'
		(Bgl II)
	R329H	5'-GCTAGATTGACT*CACTCT*CCAGTTCAAG-3' 5'-CTTGAACTGGAGA*GTGAGT*CAATCTAGC-3'

		Eco RV
	S364T	5'-CTCACGACAACACTAT*GATATC*TATTTTCTTC-3'
		5'-GAAGAAAATA*GATATC*ATAGTGTTGTCGTGAG-3'
		Nco I
	I366V	5'-CGACAACT*CCATGG*TTTCTATTTTCTTCGC-3'
		5'-GCGAAGAAAATAGAAA*CCATGG*AGTTGTCG-3'
		Kpn I
	A379K	5'-GTACAAC*GGTACC*AAGCCATTGTCTAC-3'
		5'-GTAGACAATGG**CTT***GGTACC*GTTGTAC-3'
	S396A	5'-CTGACGGTTACGCTGCTTCTTGGAC-3'
		5'-GTCCAAGAAGCAGCGTAACCGTCAG-3'
	G404A	5'-CTGTTCCATTCGCTGCTAGAGCTTAC-3'
		5'-GTAAGCTCTAGCAGCGAATGGAACAG-3'
	Q415E	5'-GATGCAATGTGAAGCTGAAAAGGAACC-3'
		5'-GGTTCCTTTTCAGCTTCACATTGCATC-3'
		Sal I
	A437G	5'-CACGGTTGTGG*TGTCGAC*AAGTTGGG-3'
		5'-CCCAACTT*GTCGAC*ACCACAACCGTG-3'
		Mun I
	A463E	5'-GATCTGGTGG*CAATTG*GGAGGAATGTTTCG-3'
		5'-CGAAACATTCCTCC*CAATTG*CCACCAGATC-3'

and accordingly for other mutations.

The temperature optimum of the purified phytases, expressed in *Saccharomyces cerevisiae* (Example 7), was determined as outlined in Example 9. Table 5 shows the effect on the stability of consensus phytase for each mutation introduced.

Table 5: Stability effect of the individual amino acid replacements in consensus phytase-1 (+ or - means a positive, respectively, negative effect on the protein stability up to 1 °C, ++ and -- means a positive, respectively, negative effect on the protein stability between 1 and 3 °C; the number 10 or 11 corresponds to the consensus phytase sequence that suggests the amino acid replacement.)

stabilizing		neutral		destabilizing	
mutation	effect	mutation	effect	mutation	effect
E58A (10)	+	D69A	±	Y54F (10)	-
D69K (11)	+	D70G (10)	±	V73I	-
D197N (10)	+	N134Q (10)	±	A94K (10)	-
T214L (10)	++	G186H	±	A101R (11)	-
E222T (11)	++	S187A (10)	±	K153N (11)	-
E267D (10)	+	T214V	±	I158V (10)	--
R291I*	+	T251N (10)	±	G203A	--
R329H (10)	+	Y259N (10)	±	G205S	-
S364T (10)	++	A283D (10)	±	A217V	-
A379K (11)	+	A320V (10)	±	V227A (11)	--
G404A (10)	++	K445T	±	L234V (10)	-
		A463E (10)	±	A238P (10)	--
				E277Q (10)	-
				H287A (11)	-
				Q292A (10)	-
				I366V (10)	-
				S396A (10)	--
				Q415E (11)	-
				A437G (10)	--
				E451R	--

*: This amino acid replacement was found in another round of mutations.

We combined eight positive mutations (E58A, D197N, E267D, R291I, R329H, S364T, A379K, G404A) in the consensus phytase using the primers and the technique mentioned above in this example. Furthermore, the mutations Q50T and K91A were introduced which mainly influence the catalytical characteristics of the phytase (see patent application EP 97810175.6 and EP 97112688 as well as Example 9). The DNA and amino acid sequence of the resulting phytase gene (consensus phytase-thermo[8]-Q50T-K91A) is shown in Figure 7. In this way, the temperature optimum and the melting point of the consensus phytase was increased by 7 °C (Figure 15, 16, 17).

Using the results of Table 5, we further improved the thermostability of consensus phytase 10 by the following back mutations K94A, V158I, and A396S that revealed a strong negative influence on the stability of consensus phytase. The resulting protein is phytase-10-thermo [3]. Furthermore, we introduced the mutations Q50T and K91A which mainly influence the catalytical characteristics of consensus phytase (see patent application EP 97810175.6 and EP 97112688 as well as Example 9 and Figure 14 and 15). The resulting DNA and amino acid sequence is shown in Figure 8. The optimized phytase showed a 4 °C higher temperature optimum and melting point than consensus phytase 10 (Figure 12 and 13). Furthermore, the phytase has also a strongly increased specific activity with phytate as substrate of 250 U/mg at pH 5.5 (Figure 14).

Example 4

Stabilization of the phytase of *A. fumigatus* ATCC 13073 by replacement of amino acid residues with the corresponding consensus phytase-1 and consensus phytase-10 residues

At six typical positions where the *A. fumigatus* 13073 is the only or nearly the only phytase in the alignment of Figure 1 that does not contain the corresponding consensus phytase amino acid residue, the non-consensus amino acid residue was replaced by the consensus one. In a first round, the following amino acids were substituted in *A. fumigatus* 13073 phytase, containing the Q27T substitution and the signal sequence of *A. terreus* cbs.116.46 phytase (see European Patent Application No. 97810175.6 and Figure 9):

F55(28)Y, V100(73)I, F114(87)Y, A243(220)L, S265(242)P, N294(282)D.

The numbers in parentheses confer to the numbering of Figure 1.

In a second round, four of the seven stabilizing amino acid exchanges (E59A, R329H, S364T, G404A) found in the consensus phytase-10 sequence and, tested as single mutation in consensus phytase-1 (Table 5), were additionally introduced into the *A. fumigatus* a-

mutant. Furthermore, the amino acid replacement S126N, shown to reduce the protease susceptibility of the phytase, was introduced.

The mutations were introduced as described in example 3 (see Table 6) and expressed as described in example 6 to 8. The resulting *A. fumigatus* 13073 phytase variants were called a-mutant and ß-mutant-E59A-S126N-R329H-S364T-G404A.

The temperature optimum (60 °C, Figure 20) and the melting point (67.0 °C, Figure 19) of the *A. fumigatus* 13073 phytase ß-mutant was increased by 5 °C in comparison to the values of the wild-type (temperature optimum: 55 °C, T_m: 60 °C). The five additional amino acid replacements further increased the temperature optimum by 3 °C (Figure 20).

Table 6: Mutagenesis primers for stabilization of *A. fumigatus* phytase ATCC 13073

Mutation	Primer
F55Y	5'-CACGTACTCGCCA**TAC**TTTTCGCTCGAG-3'
	5'-CTCGAGCGAAAAGT**ATG**GCGAGTACGTG-3'
	(*Xho* I)
E58A	5'-CCATACTTTTCG*CTCGCG*GACGAGCTGTCCGTG-3'
	5'-CACGGACAGCTCGTC*CGCGAG*CGAAAAGTAGG-3'
V100I	5'-GTATAAGAAGCTTA**TT**ACGGCGATCCAGGCC-3'
	5'-GGCCTGGATCGCCGTAA**TAA**GCTTCTTATAC-3'
F114Y	5'-CTTCAAGGGCAAG**TAC**GCCTTTTTGAAGACG-3'
	5'-CGTCTTCAAAAAGGC**GTA**CTTGCCCTTGAAG-3'
A243L	5'-CATCCGAGCTCGC**CTC**GAGAAGCATCTTC-3'
	5'-GAAGATGCTTCTC**GAG**GCGAGCTCGGATG-3'
S265P	5'-CTAATGGA TGTGT**CCG**TTTGATACGGTAG-3'
	5'-CTACCGTATCAAA**CGG**ACACATGTCCATTAG-3'

N294D 5'-GTGGAAGAAGTA**CGA**CTACCTTCAGTC-3'
5'-GACTGAAGGTAG**TCG**TACTTCTTCCAC-3'

(*Mlu* I)

R329H 5'-GCCCGGTTG*ACGCAT*TCGCCAGTGCAGG-3'
5'-CCTGCACTGGCGA*ATGCGT*CAACCGGGC-3'

Nco I

S364T 5'-CACACGACAACA*CCATGG*TTTCCATCTTC-3'
5'-GAAGATGGAAA<u>CCATGG</u>TGTTGTCGTGTG-3'

(*Bss* HI)

G404A 5'-GTGGTGCCTTTCG*CCGCGC*GAGCCTACTTC-3'
5'-GAAGTAGGCTC*GCGCGG*CGAAAGGCACCAC-3'

<u>Example 5</u>
<u>Introduction of the active site amino acid residues of the *A. niger* NRRL 3135 phytase into the consensus phytase-1</u>

We used the crystal structure of the *Aspergillus niger* NRRL 3135 phytase to define all active site amino acid residues (see Reference Example and EP 97810175.6). Using the alignment of Figure 1, we replaced the following active site residues and additionally the not identical adjacent ones of the consensus phytase by that of the *A. niger* phytase:

S89D, S92G, A94K, D164S, P201S, G203A, G205S, H212P, G224A, D226T, E255T, D256E, V258T, P265S, Q292H, G300K, Y305H, A314T, S364G, M365I, A397S, S398A, G404A, and A405S

The new protein sequence consensus phytase -7 was backtranslated into a DNA sequence (Figure 10) as described in Example 1. The corresponding gene (*fcp7*) was generated as described in Example 1 using the following oligonucleotide mixes:

Mix 1.7: CP-1, CP-2, CP-3, CP-4.7, CP-5.7, CP-6, CP-7, CP-8.7, CP-9, CP-10.7

Mix 2.7: CP-9, CP-10.7, CP-11.7, CP-12.7, CP-13.7, CP-14.7, CP-15.7, CP-16, CP-17.7, CP-18.7, CP-19.7, CP-20, CP-21, CP-22.

The DNA sequences of the oligonucleotides are indicated in Figure 3. The newly synthesized oligonucleotides are additionally marked by number 7. After assembling of the

oligonucleotides using the same PCR primers as mentioned in Example 1, the gene was cloned into an expression vector as described in Examples 6-8.

The pH-profile determined after expression in *H. polymorpha* and purification was shifted into the acidic range of the pH-spectrum showing an optimum at pH 4.5-5.0 (see Figure 18). The enzyme had a broad pH-optimum reaching at least 60% of its maximum activity from pH 2.5 to pH 6.0. Up to pH 5.0, the profile resembled the profile of the *A. niger* NRRL 3135 phytase. However, below pH 5.0 it lacked the typical low at pH 4.0 of the profile of *A. niger* phytase.

Example 6

Expression of the consensus phytase genes in *Hansenula polymorpha*

The phytase expression vectors, used to transform *H. polymorpha* RB11 (Gellissen *et al.*, 1994), was constructed by inserting the *Eco* RI fragment of pBsk⁻*fcp* or variants thereof into the multiple cloning site of the *H. polymorpha* expression vector pFPMT121, which is based on an *ura3* selection marker from *S. cerevisiae*, a formate dehydrogenase (*FMD*) promoter element and a methanol oxidase (*MO*) termimator element from *H. polymorpha*. The 5' end of the *fcp* gene is fused to the *FMD* promoter, the 3' end to the *MOX* terminator (Gellissen *et al.*, 1996; EP 0299 108 B). The resulting expression vector are designated pFPMT*fcp*, pFPMT*fcp10*, pFPMT*fcp7*.

The constructed plasmids were propagated in *E. coli*. Plasmid DNA was purified using standard state of the art procedures. The expression plasmids were transformed into the *H. polymorpha* strain RP11 deficient in orotidine-5'-phosphate decarboxylase (*ura3*) using the procedure for preparation of competent cells and for transformation of yeast as described in Gelissen *et al.* (1996). Each transformation mixture was plated on YNB (0.14% w/v Difco YNB and 0.5% ammonium sulfate) containing 2% glucose and 1.8% agar and incubated at 37 °C. After 4 to 5 days individual transformant colonies were picked and grown in the liquid medium described above for 2 days at 37 °C. Subsequently, an aliquot of this culture was used to inoculate fresh vials with YNB-medium containing 2% glucose. After seven further passages in selective medium, the expression vector integrates into the yeast genome in multimeric form. Subsequently, mitotically stable transformants were obtained by two additional cultivation steps in 3 ml non-selective liquid medium (YPD, 2% glucose, 10 g yeast extract, and 20 g peptone). In order to obtain genetically homogeneous recombinant strains an aliquot from the last stabilization culture was plated on a selective plate. Single colonies were isolated for analysis of phytase expression in YNB containing 2% glycerol instead of glucose to derepress the *fmd* promoter. Purification of the consensus phytases was done as described in Example 7.

Example 7

Expression of the consensus phytase genes in *Saccharomyces cerevisiae* and purification of the phytases from culture supernatant

The consensus phytase genes were isolated from the corresponding Bluescript-plasmid (pBsk⁻*fcp*, pBSK⁻*fcp10*, pBsk⁻*fcp7*) and ligated into the *Eco* RI sites of the expression cassette of the *Saccharomyces cerevisiae* expression vector pYES2 (Invitrogen, San Diego, CA, USA) or subcloned between the shortened GAPFL (glyceraldhyde-3-phosphate dehydrogenase) promoter and the *pho5* terminator as described by Janes *et al.* (1990). The correct orientation of the gene was checked by PCR. Transformation of *S. cerevisiae* strains. e. g. INVSc1 (Invitrogen, San Diego, CA, USA) was done according to Hinnen *et al.* (1978). Single colonies harboring the phytase gene under the control of the GAPFL promoter were picked and cultivated in 5 ml selection medium (SD-uracil, Sherman *et al.*, 1986) at 30°C under vigorous shaking (250 rpm) for one day. The preculture was then added to 500 ml YPD medium (Sherman *et al.*, 1986) and grown under the same conditions. Induction of the *gal1* promoter was done according to manufacturer's instruction. After four days of incubation cell broth was centrifuged (7000 rpm, GS3 rotor, 15 min, 5°C) to remove the cells and the supernatant was concentrated by way of ultrafiltration in Amicon 8400 cells (PM30 membranes) and ultrafree-15 centrifugal filter devices (Biomax-30K, Millipore, Bedford, MA, USA). The concentrate (10 ml) was desalted on a 40 ml Sephadex G25 Superfine column (Pharmacia Biotech, Freiburg, Germany), with 10 mM sodium acetate, pH 5.0, serving as elution buffer. The desalted sample was brought to 2 M $(NH_4)_2SO_4$ and directly loaded onto a 1 ml Butyl Sepharose 4 Fast Flow hydrophobic interaction chromatography column (Pharmacia Biotech, Feiburg, Germany) which was eluted with a linear gradient from 2 M to 0 M $(NH4)_2SO_4$ in 10 mM sodium acetate, pH 5.0. Phytase was eluted in the break-through, concentrated and loaded on a 120 ml Sephacryl S-300 gel permeation chromatography column (Pharmacia Biotech, Freiburg, Germany). Consensus phytase and consensus phytase -7 eluted as a homogeneous symmetrical peak and was shown by SDS-PAGE to be approx. 95% pure.

Example 8

Expression of the consensus phytase genes in *Aspergillus niger*

The Bluescript-plasmids pBsk⁻*fcp*, pBSK⁻*fcp10*, and pBsk⁻*fcp7* were used as template for the introduction of a *Bsp* HI-site upstream of the start codon of the genes and an *Eco* RV-site downstream of the stop codon. The Expand™ High Fidelity PCR Kit (Boehringer Mannheim, Mannheim, Germany) was used with the following primers:

Primer Asp-1:

Bsp HI

5'-TATA*TCATGA*GCGTGTTCGTCGTGCTACTGTTC-3'

Primer Asp-2 used for cloning of *fcp* and *fcp7*:

Eco RV

3'-ACCCGACTTACAAAGCGAATT*CTATAG*ATATAT-5'

Primer Asp-3 used for cloning of *fcp10*:

Eco RV

3'-ACCCTTCTTACAAAGCGAATT*CTATAG*ATATAT-5'

The reaction was performed as described by the supplier. The PCR-amplified *fcp*-genes had a new *Bsp* HI site at the start codon, introduced by primer Asp-1, which resulted in a replacement of the second amino acid residue glycine by serine. Subsequently, the DNA-fragment was digested with *Bsp* HI and *Eco* RV and ligated into the *Nco* I site downstream of the glucoamylase promoter of *Aspergillus niger* (*glaA*) and the *Eco* RV site upstream of the *Aspergillus nidulans* tryptophan C terminator (*trpC*) (Mullaney *et al.*, 1985). After this cloning step, the genes were sequenced to detect possible failures introduced by PCR. The resulting expression plasmids which basically corresponds to the pGLAC vector as described in Example 9 of EP 684 313, contained the orotidine-5'-phosphate decarboxylase gene (*pyr4*) of *Neurospora crassa* as a selection marker. Transformation of *Aspergillus niger* and expression of the consensus phytase genes was done as described in EP 684 313. The consensus phytases were purified as described in Example 7.

Example 9

Determination of phytase activity and of temperature optimum

Phytase activity was determined basically as described by Mitchell et al (1997). The activity was measured in an assay mixture containing 0.5% phytic acid (≈5 mM) in 200 mM sodium acetate, pH 5.0. After 15 min of incubation at 37 °C, the reaction was stopped by addition of an equal volume of 15% trichloroacetic acid. The liberated phosphate was quantified by mixing 100 µl of the assay mixture with 900 µl H_2O and 1 ml of 0.6 M H_2SO_4, 2% ascorbic acid and 0.5% ammonium molybdate. Standard solutions of potassium phosphate were used as reference. One unit of enzyme activity was defined as the amount of enzyme that releases 1 µmol phosphate per minute at 37 °C. The protein

concentration was determined using the enzyme extinction coefficient at 280 nm calculated according to Pace et al (1995): consensus phytase, 1.101; consensus phytase 7, 1.068; consensus phytase 10, 1.039.

In case of pH-optimum curves, purified enzymes were diluted in 10 mM sodium acetate, pH 5.0. Incubations were started by mixing aliquots of the diluted protein with an equal volume of 1% phytic acid (≈10 mM) in a series of different buffers: 0.4 M glycine/HCl, pH 2.5; 0.4 M acetate/NaOH, pH 3.0, 3.5, 4.0, 4.5, 5.0, 5.5; 0.4 M imidazole/HCl, pH 6.0, 6.5; 0.4 M Tris/HCl pH 7.0, 7.5, 8.0, 8.5, 9.0. Control experiments showed that pH was only slightly affected by the mixing step. Incubations were performed for 15 min at 37 °C as described above.

For determinations of the substrate specificities of the phytases, phytic acid in the assay mixture was replaced by 5 mM concentrations of the respective phosphate compounds. The activity tests were performed as described above.

For determination of the temperature optimum, enzyme (100 µl) and substrate solution (100 µl) were pre-incubated for 5 min at the given temperature. The reaction was started by addition of the substrate solution to the enzyme. After 15 min incubation, the reaction was stopped with trichloroacetic acid and the amount of phosphate released was determined.

The pH-optimum of the original consensus phytase was around pH 6.0-6.5 (70 U/mg). By introduction of the Q50T mutation, the pH-optimum shifted to pH 6.0 (130 U/mg). After introduction of K91A, the pH optimum shifted one pH-unit into the acidic pH-range showing a higher specific activity between pH 2.5 and pH 6.0. That was shown for the stabilized mutants and for consensus phytase-10, too (Figure 14 and 15).

Consensus phytase-7, which was constructed to transfer the catalytic characteristics of the *A. niger* phytase NRRL 3135 into the consensus phytase, had a pH-profile which is shifted into the acidic range of the pH-spectrum showing an optimum between pH 4.5 and 5.0 (see Figure 19). The enzyme had a broad pH-optimum reaching at least 60% of its increased maximum activity from pH 2.5 to pH 6.0. The substrate spectrum, too, resemble more to that of the A. niger NRRL 3135 phytase than to the consensus phytase-1.

The temperature optimum of consensus phytase-1 (71 °C) was 16-26 °C higher than the temperature optimum of the wild-type phytases (45-55 °C, Table 7) which were used to calculate the consensus sequence. The improved consensus phytase-10 showed a further

increase of its temperature optimum to 80 °C (Figure 11). The temperature optimum of the consensus phytase-1-thermo[8] was found in the same range (78 °C) using the supernatant of an overproducing *S. cerevisiae* strain. The highest temperature optimum reached of 82 ° C was determined for consensus phytase-10-thermo-Q50T-K91A.

Table 7: Temperature optimum and T_m-value of consensus phytase and of the phytases from *A. fumigatus*, *A. niger*, *E. nidulans*, and *M. thermophila*. The determination of the temperature optimum was performed as described in Example 9. The T_m-values were determined by differential scanning calorimetry as described in Example 10.

phytase	temperature optimum [°C]	*T*m [°C]
Consensus phytase-10-thermo-Q50T-K91A	82	89.3
Consensus phytase-10-thermo-Q50T	82	88.6
Consensus phytase-10	80	85.4
Consensus phytase-1-thermo[8]-Q50T	78	84.7
Consensus phytase-1-thermo[8]-Q50T-K91A	78	85.7
Consensus phytase-1	71	78.1
A. niger NRRL3135	55	63.3
A. fumigatus 13073	55	62.5
A. fumigatus 13073 α-mutant	60	67.0
A. fumigatus 13073 α-mutant (optimized)	63	-
A. terreus 9A-1	49	57.5
A. terreus cbs.116.46	45	58.5
E. nidulans	45	55.7
M. thermophila	55	n. d.
T. thermophilus	45	n. d.

Example 10

Determination of the melting point by differential scanning calorimetry (DSC)

In order to determine the unfolding temperature of the phytases, differential scanning calorimetry was applied as previously published by Brugger et al (1997). Solutions of 50-60 mg/ml homogeneous phytase were used for the tests. A constant heating rate of 10 ° C/min was applied up to 90-95 °C.

The determined melting points reflect the results obtained for the temperature optimums (Table 7). The most stable consensus phytase designed is consensus phytase-10-thermo-Q50T-K91A showing a melting temperature under the choosen condition of 89.3 ° C. This is 26 to 33.6 °C higher than the melting point of the wild-type phytases used.

Example 11

Transfer of basidiomycete phytase active site into consensus phytase-10-thermo-Q50T-K91A

As described previously (Example 3), mutations derived from the basidiomycete phytase active site were introduced into the consensus phytase 10. The following five constructs a) to e) were prepared:

a) This construct is called consensus phytase 12, and it comprises a selected number of active site residues of the basidio consensus sequence, its amino acid sequence (consphy12) is shown in Fig. 21 (the first 26 amino acids forms the signal peptide, amended positions are underlined);

b) a cluster of mutations (Cluster II) was transferred to the consensus 10 sequence, viz.: S80Q, Y86F, S90G, K91A, S92A, K93T, A94R, Y95I;

c) analogously, another cluster of mutations (Cluster III) was transferred, viz.: T129V, E133A, Q143N, M136S, V137S, N138Q, S139A;

d) analogously, a further cluster of mutations (Cluster IV) was transferred, viz.: A168D, E171T, K172N, F173W;

e) and finally, a further cluster of mutations (Cluster V) was transferred, viz.: Q297G, S298D, G300D, Y305T.

These constructs were expressed as described in Examples 6 to 8.

References:

Akanuma, S., Yamagishi, A., Tanaka, N. & Oshima, T. (1998). Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from *Bacillus subtilis* by experimental evolution. *Prot. Sci.* **7**, 698-705.

Arase, A., Yomo, T., Urabe, I., Hata, Y., Katsube, Y. & Okada, H. (1993). Stabilization of xylanase by random mutagenesis. *FEBS Lett.* **316**, 123-127.

Berka, R. M., Rey, M. W., Brown, K. M., Byun, T. & Klotz, A. V. (1998). Molecular characterization and expression of a phytase gene from the thermophilic fungus Thermomyces lanuginosus. *Appl. Environ. Microbiol.* **64**, 4423-4427.

Blaber, M., Lindstrom, J. D., Gassner, N., Xu, J., Heinz, D. W. & Matthews, B. W. (1993). Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala'Ser and Val'Thr substitutions in T4 lysozyme. *Biochemistry* **32**, 11363-11373.

Brugger, R., Mascarello, F., Augem, S., van Loon, A. P. G. M. & Wyss, M. (1997). Thermal denaturation of phytases and pH 2.5 acid phosphatase studied by differential scanning calorimetry. In *The Biochemistry of phytate and phytase (eds. Rasmussen, S.K; Raboy, V.; Dalbøge, H. and Loewus, F.;* Kluwer Academic Publishers.

Cosgrove, D.J. (1980) Inositol phosphates - their chemistry, biochemistry and physiology: studies in organic chemistry, chapter 4. Elsevier Scientific Publishing Company, Amsterdam, Oxford, New York.

Devereux, J., Haeberli, P.& Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387-395.

Gellissen, G., Hollenberg, C. P., Janowicz, Z. A. (1994) Gene expression in methylotrophic yeasts . In: Smith, A. (ed.) Gene expression in recombinant microorganisms. Dekker, New York, 395-439.

Gellissen, G., Piontek, M., Dahlems, U., Jenzelewski, V., Gavagan, J. E., DiCosimo, R., Anton, D. I. & Janowicz, Z. A. (1996) Recombinant *Hansenula polymorpha* as a biocatalyst: coexpression of the spinach glycolate oxidase (*GO*) and the *S. cerevisiae* catalase T (*CTT1*) gene. *Appl. Microbiol. Biotechnol.* **46**, 46-54.

Gerber, P. and Müller, K. (1995) Moloc molecular modeling software. *J. Comput. Aided Mol. Des.* **9**, 251-268

Hinnen, A., Hicks, J. B. & Fink, G, R. (1978) Transformation of yeast. *Proc. Natl. Acad. Sci. USA* **75**, 1929-1933.

Imanaka, T., Shibazaki, M. & Takagi, M. (1986). A new way of enhancing the thermostability of proteases. *Nature* **324**, 695-697.

Janes, M., Meyhack, B., Zimmermann, W. & Hinnen, A. (1990) The influence of GAP promoter variants on hirudine production, average plasmid copy number and cell growth in *Saccharomyces cerevisiae*. *Curr. Genet.* **18**, 97-103.

Karpusas, M., Baase, W. A., Matsumura, M. & Matthews, B. W. (1989). Hydrophobic packing in T4 lysozyme probed by cavity-filling mutants. *Proc. Natl. Acad. Sci.(USA)* **86**, 8237-8241.

Margarit, I., Campagnoli, S., Frigerio, F., Grandi, G., Fillipis, V. D. & Fontana, A. (1992). Cumulative stabilizing effects of glycine to alanine substitutions in *Bacillus subtilis* neutral protease. *Prot. Eng.* **5**, 543-550.

Matthews, B. W. (1987a). Genetic and structural analysis of the protein stability problem. *Biochemistry* **26**, 6885-6888.

Matthews, B. W. (1993). Structural and genetic analysis of protein stability. *Annu. Rev. Biochem.* **62**, 139-160.

Matthews, B. W., Nicholson, H. & Becktel, W. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. (USA)* **84**, 6663-6667.

Mitchell, D. B., Vogel, K., Weimann, B. J., Pasamontes, L. & van Loon, A. P. G. M. (1997) The phytase subfamily of histidine acid phosphatases: isolation of genes for two novel phytases from the fungi *Aspergillus terreus* and *Myceliophthora thermophila*, *Microbiology* **143**, 245-252.

Mullaney, E. J., Hamer, J. E., Roberti, K. A., Yelton, M. M. & Timberlake, W. E. (1985) Primary structure of the *trpC* gene from *Aspergillus nidulans*. *Mol. Gen. Genet.* **199**, 37-46.

Munoz, V. & Serrano, L. (1995). Helix design, prediction and stability. *Curr. Opin. Biotechnol.* **6**, 382-386.

Pace, N. C., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Prot. Sci.* **4**, 2411-2423.

Pantoliano, M. W., Landner, R. C., Brian, P. N., Rollence, M. L., Wood, J. F. & Poulos, T. L. (1987). Protein engineering of subtilisin BPN': enhanced stabilization through the introduction of two cysteines to form a disulfide bond. *Biochemistry* **26**, 2077-2082.

Pasamontes, L., Haiker, M., Henriquez-Huecas, M., Mitchell, D. B. & van Loon, A. P. G. M. (1997a). Cloning of the phytases from *Emericella nidulans* and the thermophilic fungus *Talaromyces thermophilus*. *Biochim. Biophys. Acta* **1353**, 217-223.

Pasamontes, L., Haiker, M., Wyss, M., Tessier, M. & van Loon, A. P. G. M. (1997) Cloning, purification and characterization of a heat stable phytase from the fungus *Aspergillus fumigatus*, *Appl. Environ. Microbiol.* **63**, 1696-1700.

Piddington, C. S., Houston, C. S., Paloheimo, M., Cantrell, M., Miettinen-Oinonen, A. Nevalainen, H., & Rambosek, J. (1993) The cloning and sequencing of the genes encoding phytase (*phy*) and pH 2.5-optimum acid phosphatase (*aph*) from *Aspergillus niger* var. *awamori*. *Gene* **133**, 55-62.

Purvis, I. J., Bettany, A. J. E., Santiago, T. C., Coggins, J. R., Duncan, K., Eason, R. & Brown, A. J. P. (1987). The efficiency of folding of some proteins is increased by controlled rates of translation *in vivo*. *J. Mol. Biol.* **193**, 413-417.

Risse, B., Stempfer, G., Rudolph, R., Schumacher, G. & Jaenicke, R. (1992). Characterization of the stability effect of point mutations of pyruvate oxidase from *Lactobacillus plantarum*: protection of the native state by modulating coenzyme binding and subunit interaction. *Prot. Sci.* **1**, 1710-1718.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sauer, R., Hehir, K., Stearman, R., Weiss, M., Jeitler-Nilsson, A., Suchanek, E. & Pabo, C. (1986). An engineered intersubunit disulfide enhances the stability and DNA binding of the N-terminal domain of l-repressor. *Biochemistry* **25**, 5992-5999.

Serrano, L., Day, A. G. & Fersht, A. R. (1993). Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J. Mol. Biol.* **233**, 305-312.

Sheman, J. P., Finck, G. R. & Hicks, J. B. (1986) Laboratory course manual for methods in yeast genetics. Cold Spring Harbor University.

Steipe, B., Schiller, B., Plueckthun, A. & Steinbach, S. (1994). Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* **240**, 188-192.

van den Burg, B., Vriend, G., Veltman, O. R., Venema & G., Eijsink, V. G. H. (1998). Engineering an enzyme to resist boiling. *Proc. Natl. Acad. Sci. (USA)* **95**, 2056-2060.

Van Etten, R.L. (1982) Human prostatic acid phosphatase: a histidine phosphatase. *Ann. NY Acad. Sci.* **390**,27-50

van Hartingsveldt, W., van Zeijl, C. M. F., Harteveld, G. M., Gouka, R. J., Suykerbuyk, M. E. G., Luiten, R. G. M., van Paridon, P. A., Selten, G. C. M., Veenstra, A. E., van Gorcom, R. F. M., & van den Hondel, C. A. M. J. J. (1993) Cloning, characterization and overexpression of the phytase-encoding gene (*phyA*) of *Aspergillus niger*. *Gene* **127**, 87-94.

van Loon, A. P. G. M., Simoes-Nunes, C., Wyss, M., Tomschy, A., Hug, D., Vogel, K. & Pasamontes, L. (1998). A heat resistant phytase of *Aspergillus fumigatus* with superior performance in animal experiments. Phytase optimization and natural variability. In *the Biochemistry of phytate and phytases*. Kluwer Academic Press, s.a.

Claims

1. A process for the preparation of a consensus protein, whereby such process is characterized by the following steps:

a) at least three, preferably four amino acid sequences are aligned by any standard alignment program known in the art,

b) amino acids at the same position according to such alignment are compared regarding their evolutionary similarity by any standard program known in the art, whereas the degree of similarity provided by such a program which defines the least similarity of the amino acids that is used for the determination of an amino acid of corresponding positions is set to a less stringent number and the parameters are set in such a way that it is possible for the program to determine from only 2 identical amino acids at a corresponding position an amino acid for the consensus protein; however, if among the compared amino acid sequences are sequences that show a much higher degree of similarity to each other than to the residual sequences, these sequences are represented by their consensus sequence determined as defined in the same way as in the present process for the consensus sequence of the consensus protein or a vote weight of 1 divided by the number of such sequences is assigned to every of -those sequences,

c) in case no common amino acid at a defined position is identified by the program, any of the amino acids, preferably the most frequent amino acid of all such sequences is selected,

d) once the consensus sequence has been defined, such sequence is back-translated into a DNA sequence, preferably by using a codon frequency table of the organism in which expression should take place,

e) the DNA sequence is synthesized by methods known in the art and used either integrated into a suitable expression vector or by itself to transform an appropriate host cell,

f) the transformed host cell is grown under suitable culture conditions and the consensus protein is isolated from the host cell or its culture medium by methods known in the art.

2. A process as claimed in claim 1 wherein the program used for the comparison of amino acids at a defined position regarding their evolutionary similarity is the program "PRETTY".

3. A process as claimed in claims 1 or 2, wherein

in a first step a consensus sequence is determined from a number of highly homologous sequences according to steps a), b) and c) of claim 1,

in a second step the amino acid sequence of another protein which is homologous to the consensus sequence is compared with the consensus sequence and

in a third step only those amino acid residues are replaced in the amino acid sequence of the other protein which clearly differ from the consensus sequence of this protein family calculated under moderately stringent conditions whereas at all positions of the alignment where no preferred single amino acid can be determined under moderately stringent conditions the amino acids of the other protein remain unchanged.

4. A process as claimed in any one of claims 1-3, wherein

in a first step a consensus sequence is determined from homologous sequences according to steps a), b) and c) of claim 1,

in a second step the active center of the protein comprising all amino acid residues that are involved in forming the active center is determined in the consensus sequence and in the sequence of a homologous protein as well and

in a third step some or all of the amino acids that form the active center of the homologeous protein are inserted in the backbone of the consensus sequence.

5. A process as claimed in claim 4, wherein the active center of the protein is determined by using an analysis of the three-dimensional structure of the protein.

6. A process as claimed in claims 4 and 5, wherein the homologeous protein is an enzyme.

7. A process as claimed in claims 1 to *6*, wherein the defined protein family is the family of phytases.

8. A process as claimed in claim *7*, wherein the phytases are of fungal origin.

9. A process as claimed in claims 7 or 8, wherein the amino acid sequence of the phytase is changed by the introduction of at least one mutation selected from the group consisting of

E58A	F54Y
D69K	I73V
D197N	K94A
T214L	R101A
E222T	N153K
E267D	V158I
R291I	A203G
R329H	S205G
S364T	V217A
A379K	A227V
G404A	V234L
	P238A
	Q277E
	A287H
	A292Q
	V366I
	A396S
	E415Q
	G437A
	E451R

whereby the number represents the position in the consensus phytase sequence or a corresponding residue according to an alignment as shown in Fig. 1 when 26 amino acids (signal sequence) are added to the sequences shown in Fig. 1 and the letter before the number represents the amimo acid in the phytase which is replaced by the amino acid behind the number.

10. A process as claimed in any one of claims 1 to 9, wherein the host cell is of eukaryotic origin.

11. A process as claimed in claim 10, wherein eukaryotic means fungal, preferably Aspergillus or yeast, preferably Saccharomyces or Hansenula.

12. A consensus protein obtainable preferably obtained by a process as claimed in any one of claims 1 to 11.

13. A consensus protein which comprises the amino acid sequence shown in Figure 2 or any variants or muteins thereof (consensus phytase-1).

14. A mutein of the consensus protein of claim 13 characterized therein that in the amino acid sequence of Figure 2 the following replacements have been effected Q50L, Q50T, Q50G, Q50T-Y51N, Q50L-Y51N or Q50T-K91A.

15. A consensus protein which comprises the amino acid sequence shown in Figure 4 having the designation consensus phytase 10 (Fcp10) and any variants or muteins thereof.

16. A consensus protein which comprises the amino acid sequence shown in Figure 6 having the designation Consensus seq. 11 and any variant or mutein thereof.

17. A consensus protein which comprises the amino acid sequence shown in Figure 10 (consensus phytase 7) and any variant or mutein thereof.

18. A consensus protein which comprises the amino acid sequence shown in Figure 21 (consensus phytase 12) and any variant or mutein thereof.

19. A consensus protein which comprises the amino acid sequence shown in Figure 3 (basidio consensus) and any variant or mutein thereof.

20. A phytase being selected from amongst: A. fumigatus ATCC 13073 alpha-mutant; A. fumigatus ATCC 13073 alpha-mutant-(E59A-S126N-R329H, S364T-G404A)-Q27T; A. fumigatus ATCC 13073 alpha-mutant-(E59A, S126N-R329H-S364T-G404A)-Q27T-K68A, preferably the latter.

21. A food, feed or pharmaceutical composition comprising a consensus protein as claimed in any of the claims 12 to 17.

Figure 1

```
                                1
    50
 5  A. terreus 9A-1            KhsDCNSVDh GYQCFPELSH kWGlYAPYFS
    LQDESPFPlD VPEDChITFV
    A. terreus cbs             NhsDCTSVDr GYQCFPELSH kWGlYAPYFS
    LQDESPFPlD VPDDChITFV
    A. niger var. awamori      NqsTCDTVDQ GYQCFSETSH LWGQYAPFFS
10  LANESAISPD VPAGCrVTFA
    A. niger T213              NqsSCDTVDQ GYQCFSETSH LWGQYAPFFS
    LANESVISPD VPAGCrVTFA
    A. niger NRRL3135          NqsSCDTVDQ GYQCFSETSH LWGQYAPFFS
    LANESVISPE VPAGCrVTFA
15  A. fumigatus 13073         GSkSCDTVDl GYQCsPATSH LWGQYSPFFS
    LEDElSVSSK LPKDCrITLV
    A. fumigatus 32722         GSkSCDTVDl GYQCsPATSH LWGQYSPFFS
    LEDElSVSSK LPKDCrITLV
    A. fumigatus 58128         GSkSCDTVDl GYQCsPATSH LWGQYSPFFS
20  LEDElSVSSK LPKDCrITLV
    A. fumigatus 26906         GSkSCDTVDl GYQCsPATSH LWGQYSPFFS
    LEDElSVSSK LPKDCrITLV
    A. fumigatus 32239         GSkACDTVEl GYQCsPGTSH LWGQYSPFFS
    LEDElSVSSD LPKDCrVTFV
25  E. nidulans                QNHSCNTADG GYQCFPNVSH VWGQYSPYFS
    IEQESAISeD VPHGCeVTFV
    T. thermophilus            DSHSCNTVEG GYQCrPEISH sWGQYSPFFS
    LADQSEISPD VPQNCkITFV
    M. thermophila             ESRPCDTpDl GFQCgTAISH FWGQYSPYFS
30  VpSElDaS.. IPDDCeVTFA

    Consensus                  NSHSCDTVDG GYQCFPEISH LWGQYSPYFS
    LEDESAISPD VPDDC-VTFV
    Consensus phytase          NSHSCDTVDG GYQCFPEISH LWGQYSPYFS
35  LEDESAISPD VPDDCRVTFV


                                51
    100
40  A. terreus 9A-1            QVLARHGARs PThSKtKAYA AtIAAIQKSA
    TaFpGKYAFL QSYNYSLDSE
    A. terreus cbs             QVLARHGARs PTDSKtKAYA AtIAAIQKNA
    TaLpGKYAFL KSYNYSMGSE
    A. niger var. awamori      QVLSRHGARY PTESKgKkYS ALIEEIQQNV
45  TtFDGKYAFL KTYNYSLGAD
    A. niger T213              QVLSRHGARY PTESKgKkYS ALIEEIQQNV
    TtFDGKYAFL KTYNYSLGAD
    A. niger NRRL3135          QVLSRHGARY PTDSKgKkYS ALIEEIQQNA
    TtFDGKYAFL KTYNYSLGAD
50  A. fumigatus 13073         QVLSRHGARY PTSSKsKkYK kLVTAIQaNA
    TdFKGKFAFL KTYNYTLGAD
    A. fumigatus 32722         QVLSRHGARY PTSSKsKkYK kLVTAIQaNA
    TdFKGKFAFL KTYNYTLGAD
    A. fumigatus 58128         QVLSRHGARY PTSSKsKkYK kLVTAIQaNA
55  TdFKGKFAFL KTYNYTLGAD
```

A. fumigatus 26906 QVLSRHGARY PTSSKsKkYK kLVTAIQaNA TdFKGKFAFL KTYNYTLGAD
A. fumigatus 32239 QVLSRHGARY PTASKsKkYK kLVTAIQKNA TeFKGKFAFL ETYNYTLGAD
E. nidulans QVLSRHGARY PTESKsKAYS GLIEAIQKNA TsFwGQYAFL ESYNYTLGAD
T. thermophilus QLLSRHGARY PTSSKtElYS QLISrIQKTA TaYKGyYAFL KDYrYqLGAN
M. thermophila QVLSRHGARa PTlKRaaSYv DLIDrIHhGA IsYgPgYEFL RTYDYTLGAD

Consensus QVLSRHGARY PTSSK-KAYS ALIEAIQKNA T-FKGKYAFL KTYNYTLGAD
Consensus phytase QVLSRHGARY PTSSKSKAYS ALIEAIQKNA TAFKGKYAFL KTYNYTLGAD

101 150

A. terreus 9A-1 ELTPFGrNQL rDlGaQFYeR YNALTRhInP FVRATDASRV hESAEKFVEG
A. terreus cbs NLTPFGrNQL qDlGaQFYRR YDTLTRhInP FVRAADSSRV hESAEKFVEG
A. niger var. *awamori* DLTPFGEQEL VNSGIKFYQR YESLTRNIIP FIRSSGSSRV IASGEKFIEG
A. niger T213 DLTPFGEQEL VNSGIKFYQR YESLTRNIIP FIRSSGSSRV IASGEKFIEG
A. niger NRRL3135 DLTPFGEQEL VNSGIKFYQR YESLTRNIVP FIRSSGSSRV IASGKKFIEG
A. fumigatus 13073 DLTPFGEQQL VNSGIKFYQR YKALARSVVP FIRASGSDRV IASGEKFIEG
A. fumigatus 32722 DLTPFGEQQL VNSGIKFYQR YKALARSVVP FIRASGSDRV IASGEKFIEG
A. fumigatus 58128 DLTPFGEQQL VNSGIKFYQR YKALARSVVP FIRASGSDRV IASGEKFIEG
A. fumigatus 26906 DLTAFGEQQL VNSGIKFYQR YKALARSVVP FIRASGSDRV IASGEKFIEG
A. fumigatus 32239 DLTPFGEQQM VNSGIKFYQK YKALAgSVVP FIRSSGSDRV IASGEKFIEG
E. nidulans DLTiFGENQM VDSGaKFYRR YKNLARKnTP FIRASGSDRV VASAEKFING
T. thermophilus DLTPFGENQM IQlGIKFYnH YKSLARNaVP FVRCSGSDRV IASGrlFIEG
M. thermophila ELTRtGQQQM VNSGIKFYRR YRALARKsIP FVRTAGqDRV VhSAENFTQG

Consensus DLTPFGENQM VNSGIKFYRR YKALARK-VP FVRASGSDRV IASAEKFIEG
Consensus phytase DLTPFGENQM VNSGIKFYRR YKALARKIVP FIRASGSDRV IASAEKFIEG

```
                                151
      200
      A. terreus 9A-1           FQTARqDDHh ANpHQPSPrV DVaIPEGSAY
      NNTLEHSlCT AFES...STV
      A. terreus cbs            FQNARqGDPh ANpHQPSPrV DVVIPEGTAY
      NNTLEHSICT AFEA...STV
      A. niger var. awamori     FQSTKLkDPr AqpgQSSPkI DVVISEASSs
      NNTLDPGTCT VFED...SEL
      A. niger T213             FQSTKLkDPr AqpgQSSPkI DVVISEASSs
      NNTLDPGTCT VFED...SEL
      A. niger NRRL3135         FQSTKLkDPr AqpgQSSPkI DVVISEASSs
      NNTLDPGTCT VFED...SEL
      A. fumigatus 13073        FQqAKLADPG A.TNRAAPAI SVIIPESETF
      NNTLDHGVCT kFEA...SQL
      A. fumigatus 32722        FQqAKLADPG A.TNRAAPAI SVIIPESETF
      NNTLDHGVCT kFEA...SQL
      A. fumigatus 58128        FQqAKLADPG A.TNRAAPAI SVIIPESETF
      NNTLDHGVCT kFEA...SQL
      A. fumigatus 26906        FQqAKLADPG A.TNRAAPAI SVIIPESETF
      NNTLDHGVCT kFEA...SQL
      A. fumigatus 32239        FQqANVADPG A.TNRAAPVI SVIIPESETY
      NNTLDHSVCT NFEA...SEL
      E. nidulans               FRKAQLhDHG S..gQATPVV NVIIPEiDGF
      NNTLDHSTCV SFEN...DEr
      T. thermophilus           FQSAKVlDPh SDkHDAPPTI NVIIeEGPSY
      NNTLDtGSCP VFED...SSg
      M. thermophila            FHSAlLADRG STvRPTlPyd mVVIPETAGa
      NNTLHNDlCT AFEEgpySTI

      Consensus                 FQSAKLADPG S-PHQASPVI NVIIPEGSGY
      NNTLDHGTCT AFED---SEL
      Consensus phytase         FQSAKLADPG SQPHQASPVI DVIIPEGSGY
      NNTLDHGTCT AFED...SEL

                                201
      250
      A. terreus 9A-1           GDDAvANFTA VFAPAIaQRL EADLPGVqLS
      TDDVVnLMAM CPFETVSlTD
      A. terreus cbs            GDAAADNFTA VFAPAIakRL EADLPGVqLS
      ADDVVnLMAM CPFETVSlTD
      A. niger var. awamori     ADTVEANFTA TFAPSIRQRL ENDLSGVTLT
      DTEVTyLMDM CSFDTIStST
      A. niger T213             ADTVEANFTA TFAPSIRQRL ENDLSGVTLT
      DTEVTyLMDM CSFDTIStST
      A. niger NRRL3135         ADTVEANFTA TFVPSIRQRL ENDLSGVTLT
      DTEVTyLMDM CSFDTIStST
      A. fumigatus 13073        GDEVAANFTA lFAPDIRARa EkHLPGVTLT
      DEDVVsLMDM CSFDTVARTS
      A. fumigatus 32722        GDEVAANFTA lFAPDIRARa EkHLPGVTLT
      DEDVVsLMDM CSFDTVARTS
      A. fumigatus 58128        GDEVAANFTA lFAPDIRARa EkHLPGVTLT
      DEDVVsLMDM CSFDTVARTS
      A. fumigatus 26906        GDEVAANFTA lFAPDIRARa KkHLPGVTLT
      DEDVVsLMDM CSFDTVARTS
      A. fumigatus 32239        GDEVEANFTA lFAPAIRARI EkHLPGVqLT
      DDDVVsLMDM CSFDTVARTA
```

```
     E. nidulans              ADEiEANFTA IMGPPIRkRL ENDLPGIKLT
     NENVIyLMDM CSFDTMARTA
     T. thermophilus          GHDAQEKFAk qFAPAIlEKI KDHLPGVDLA
     vSDVpyLMDL CPFETLARNh
     M. thermophila           GDDAQDTYlS TFAGPItARV NANLPGANLT
     DADTVaLMDL CPFETVAsSS

     Consensus                GDDAEANFTA TFAPAIRARL EADLPGVTLT DEDVV-
     LMDM CPFETVARTS
     Consensus phytase        GDDVEANFTA LFAPAIRARL EADLPGVTLT
     DEDVVYLMDM CPFETVARTS


                              251
     300
     A. terreus 9A-1          ........ ...DAhTLSPFC DLFTAtEWtq
     YNYLlSLDKY YGYGGGNPLG
     A. terreus cbs           ........ ...DAhTLSPFC DLFTAaEWtq
     YNYLlSLDKY YGYGGGNPLG
     A. niger var. awamori    ........ ...vDTKLSPFC DLFTHdEWih
     YDYLQSLkKY YGHGAGNPLG
     A. niger T213            ........ ...vDTKLSPFC DLFTHdEWih
     YDYLRSLkKY YGHGAGNPLG
     A. niger NRRL3135        ........ ...vDTKLSPFC DLFTHdEWin
     YDYLQSLkKY YGHGAGNPLG
     A. fumigatus 13073       ........ ...DASQLSPFC QLFTHnEWkk
     YNYLQSLGKY YGYGAGNPLG
     A. fumigatus 32722       ........ ...DASQLSPFC QLFTHnEWkk
     YNYLQSLGKY YGYGAGNPLG
     A. fumigatus 58128       ........ ...DASQLSPFC QLFTHnEWkk
     YNYLQSLGKY YGYGAGNPLG
     A. fumigatus 26906       ........ ...DASQLSPFC QLFTHnEWkk
     YNYLQSLGKY YGYGAGNPLG
     A. fumigatus 32239       ........ ...DASELSPFC AIFTHnEWkk
     YDYLQSLGKY YGYGAGNPLG
     E. nidulans              ........ ...HGTELSPFC AIFTEkEWlq
     YDYLQSLSKY YGYGAGSPLG
     T. thermophilus          ........ ...TDT.LSPFC ALsTQeEWqa
     YDYYQSLGKY YGnGGGNPLG
     M. thermophila           sdpatadagg gNGrpLSPFC rLFSEsEWra
     YDYLQSVGKW YGYGPGNPLG

     Consensus                ---------- -DATELSPFC ALFTE-EW--
     YDYLQSLGKY YGYGAGNPLG
     Consensus phytase        .......... .DATELSPFC ALFTHDEWRQ
     YDYLQSLGKY YGYGAGNPLG
```

```
                                301
      350
      A. terreus 9A-1           PVQGVGWaNE LMARLTRAPV HDHTCVNNTL
      DASPATFPLN ATLYADFSHD
      A. terreus cbs            PVQGVGWaNE LIARLTRSPV HDHTCVNNTL
      DANPATFPLN ATLYADFSHD
      A. niger var. awamori     PTQGVGYaNE LIARLTHSPV HDDTSSNHTL
      DSNPATFPLN STLYADFSHD
      A. niger T213             PTQGVGYaNE LIARLTHSPV HDDTSSNHTL
      DSNPATFPLN STLYADFSHD
      A. niger NRRL3135         PTQGVGYaNE LIARLTHSPV HDDTSSNHTL
      DSSPATFPLN STLYADFSHD
      A. fumigatus 13073        PAQGIGFtNE LIARLTRSPV QDHTSTNsTL
      vSNPATFPLN ATMYVDFSHD
      A. fumigatus 32722        PAQGIGFtNE LIARLTRSPV QDHTSTNsTL
      vSNPATFPLN ATMYVDFSHD
      A. fumigatus 58128        PAQGIGFtNE LIARLTRSPV QDHTSTNsTL
      vSNPATFPLN ATMYVDFSHD
      A. fumigatus 26906        PAQGIGFtNE LIARLTRSPV QDHTSTNsTL
      vSNPATFPLN ATMYVDFSHD
      A. fumigatus 32239        PAQGIGFtNE LIARLTNSPV QDHTSTNsTL
      DSDPATFPLN ATIYVDFSHD
      E. nidulans               PAQGIGFtNE LIARLTQSPV QDNTSTNHTL
      DSNPATFPLD rKLYADFSHD
      T. thermophilus           PAQGVGFvNE LIARMTHSPV QDYTTVNHTL
      DSNPATFPLN ATLYADFSHD
      M. thermophila            PTQGVGFvNE LLARLAgvPV RDgTSTNRTL
      DGDPrTFPLG rPLYADFSHD

      Consensus                 PAQGVGF-NE LIARLTHSPV QDHTSTNHTL
      DSNPATFPLN ATLYADFSHD
      Consensus phytase         PAQGVGFANE LIARLTRSPV QDHTSTNHTL
      DSNPATFPLN ATLYADFSHD

                                351
      400
      A. terreus 9A-1           SNLVSIFWAL GLYNGTAPLS qTSVESVSQT
      DGYAAAWTVP FAARAYVEMM
      A. terreus cbs            SNLVSIFWAL GLYNGTkPLS qTTVEDITrT
      DGYAAAWTVP FAARAYIEMM
      A. niger var. awamori     NGIISILFAL GLYNGTkPLS TTTVENITQT
      DGFSSAWTVP FASRlYVEMM
      A. niger T213             NGIISILFAL GLYNGTkPLS TTTVENITQT
      DGFSSAWTVP FASRlYVEMM
      A. niger NRRL3135         NGIISILFAL GLYNGTkPLS TTTVENITQT
      DGFSSAWTVP FASRlYVEMM
      A. fumigatus 13073        NSMVSIFFAL GLYNGTEPLS rTSVESaKEl
      DGYSASWVVP FGARAYFEtM
      A. fumigatus 32722        NSMVSIFFAL GLYNGTGPLS rTSVESaKEl
      DGYSASWVVP FGARAYFEtM
      A. fumigatus 58128        NSMVSIFFAL GLYNGTEPLS rTSVESaKEl
      DGYSASWVVP FGARAYFEtM
      A. fumigatus 26906        NSMVSIFFAL GLYNGTEPLS rTSVESaKEl
      DGYSASWVVP FGARAYFEtM
      A. fumigatus 32239        NGMIPIFFAM GLYNGTEPLS qTSeESTKES
      NGYSASWAVP FGARAYFEtM
```

E. nidulans NSMISIFFAM GLYNGTQPLS mDSVESIQEm
DGYAASWTVP FGARAYFELM
T. thermophilus NTMTSIFaAL GLYNGTAkLS TTEIKSIEET
DGYSAAWTVP FGGRAYIEMM
M. thermophila NDMMGVLgAL GaYDGVPPLD KTArrDpEEl
GGYAASWAVP FAARiYVEKM

Consensus NSMISIFFAL GLYNGTAPLS TTSVESIEET
DGYAASWTVP FGARAYVEMM
Consensus phytase NSMISIFFAL GLYNGTAPLS TTSVESIEET
DGYSASWTVP FGARAYVEMM

401

450
A. terreus 9A-1 QC..... .....RAEKE PLVRVLVNDR
VMPLHGCPTD KLGRCKrDAF
A. terreus cbs QC..... .....RAEKQ PLVRVLVNDR
VMPLHGCAVD NLGRCKrDDF
A. niger var. *awamori* QC..... .....QAEQE PLVRVLVNDR
VVPLHGCPID aLGRCTrDSF
A. niger T213 QC..... .....QAEQE PLVRVLVNDR
VVPLHGCPID aLGRCTrDSF
A. niger NRRL3135 QC..... .....QAEQE PLVRVLVNDR
VVPLHGCPVD aLGRCTrDSF
A. fumigatus 13073 QC..... .....KSEKE PLVRALINDR
VVPLHGCDVD KLGRCKLNDF
A. fumigatus 32722 QC..... .....KSEKE PLVRALINDR
VVPLHGCDVD KLGRCKLNDF
A. fumigatus 58128 QC..... .....KSEKE SLVRALINDR
VVPLHGCDVD KLGRCKLNDF
A. fumigatus 26906 QC..... .....KSEKE PLVRALINDR
VVPLHGCDVD KLGRCKLNDF
A. fumigatus 32239 QC..... .....KSEKE PLVRALINDR
VVPLHGCAVD KLGRCKLKDF
E. nidulans QC..... .....E.KKE PLVRVLVNDR
VVPLHGCAVD KFGRCTLDDW
T. thermophilus QC..... .....DDSDE PVVRVLVNDR
VVPLHGCEVD SLGRCKrDDF
M. thermophila RCsgggggggg ggegrQEKDE eMVRVLVNDR
VMTLkGCGAD ErGMCTLErF

Consensus QC-------- -----QAEKE PLVRVLVNDR
VVPLHGCAVD KLGRCKLDDF
Consensus phytase QC..... .....QAEKE PLVRVLVNDR
VVPLHGCAVD KLGRCKRDDF

451

471
A. terreus 9A-1 VAGLSFAQAG GNWADCF~~~ ~
A. terreus cbs VEGLSFARAG
GNWAECF~~~ ~
A. niger var. *awamori* VrGLSFARSG GDWAECsA~~ ~
A. niger T213 VrGLSFARSG GDWAECFA~~ ~
A. niger NRRL3135 VrGLSFARSG
GDWAECFA~~ ~
A. fumigatus 13073 VKGLSWARSG GNWGECFS~~ ~
A. fumigatus 32722 VKGLSWARSG GNWGECFS~~ ~

```
    A. fumigatus 58128     VKGLSWARSG GNWGECFS~~ ~
    A. fumigatus 26906     VKGLSWARSG GNWGECFS~~ ~
    A. fumigatus 32239     VKGLSWARSG
    GNSEQSFS~~ ~
 5  E. nidulans            VEGLNFARSG GNWkTCFTl~ ~
    T. thermophilus        VrGLSFARqG GNWEGCYAas e
    M. thermophila         IESMAFARGN GKWDlCFA~~ ~

    Consensus              VEGLSFARSG GNWAECFA-- -
10  Consensus phytase      VEGLSFARSG GNWAECFA.. .
```

Figure 2

```
                CP-1

                   Eco RI  M  G  V  F  V  V  L  L  S  I  A  T  L  F  G  S  T
15              TATATGAATTCATGGGCGTGTTCGTCGTGCTACTGTCCATTGCCACCTTGTTCGGTTCCA

              1 ---------+---------+---------+---------+---------+---------+ 60

                ATATACTTAAGTACCCGCACAAGCAGCACGATGACAGGTAACGGTGGAACAAGCCAAGGT

                  S  G  T  A  L  G  P  R  G  N  S  H  S  C  D  T  V  D  G  G

20              CATCCGGTACCGCCTTGGGTCCTCGTGGTAATTCTCACTCTTGTGACACTGTTGACGGTG

             61 ---------+---------+---------+---------+---------+---------+ 120

                GTAGGCCATGGCGGAACCCAGGAGCACCATTAAGAGTGAGAACACTGTGACAACTGCCAC

                     CP-2
                          CP-3

25                Y  Q  C  F  P  E  I  S  H  L  W  G  Q  Y  S  P  Y  F  S  L
                  GTTACCAATGTTTCCCAGAAATTTCTCACTTGTGGGGTCAATACTCTCCATACTTCTCTT

            121 ---------+---------+---------+---------+---------+---------+ 180

                CAATGGTTACAAAGGGTCTTTAAAGAGTGAACACCCCAGTTATGAGAGGTATGAAGAGAA

                  E  D  E  S  A  I  S  P  D  V  P  D  D  C  R  V  T  F  V  Q
30
                TGGAAGACGAATCTGCTATTTCTCCAGACGTTCCAGACGACTGTAGAGTTACTTTCGTTC

            181 ---------+---------+---------+---------+---------+---------+ 240

                ACCTTCTGCTTAGACGATAAAGAGGTCTGCAAGGTCTGCTGACATCTCAATGAAAGCAAG

                           CP-4
35                              CP-5

                  V  L  S  R  H  G  A  R  Y  P  T  S  S  K  S  K  A  Y  S  A

                AAGTTTTGTCTAGACACGGTGCTAGATACCCAACTTCTTCTAAGTCTAAGGCTTACTCTG
```

```
241 ---------+---------+---------+---------+---------+---------+ 300
    TTCAAAACAGATCTGTGCCACGATCTATGGGTTGAAGAAGATTCAGATTCCGAATGAGAC
      L  I  E  A  I  Q  K  N  A  T  A  F  K  G  K  Y  A  F  L  K
    CTTTGATTGAAGCTATTCAAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGA
301 ---------+---------+---------+---------+---------+---------+ 360
    GAAACTAACTTCGATAAGTTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACT
                                CP-6
                                  CP-7
      T  Y  N  Y  T  L  G  A  D  D  L  T  P  F  G  E  N  Q  M  V
    AGACTTACAACTACACTTTGGGTGCTGACGACTTGACTCCATTCGGTGAAAACCAAATGG
361 ---------+---------+---------+---------+---------+---------+ 420
    TCTGAATGTTGATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTTTGGTTTACC
      N  S  G  I  K  F  Y  R  R  Y  K  A  L  A  R  K  I  V  P  F
    TTAACTCTGGTATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCAT
421 ---------+---------+---------+---------+---------+---------+ 480
    AATTGAGACCATAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTA
                                CP-8
                                  CP-9
      I  R  A  S  G  S  D  R  V  I  A  S  A  E  K  F  I  E  G  F
    TCATTAGAGCTTCTGGTTCTGACAGAGTTATTGCTTCTGCTGAAAAGTTCATTGAAGGTT
481 ---------+---------+---------+---------+---------+---------+ 540
    AGTAATCTCGAAGACCAAGACTGTCTCAATAACGAAGACGACTTTTCAAGTAACTTCCAA
      Q  S  A  K  L  A  D  P  G  S  Q  P  H  Q  A  S  P  V  I  D
    TCCAATCTGCTAAGTTGGCTGACCCAGGTTCTCAACCACACCAAGCTTCTCCAGTTATTG
541 ---------+---------+---------+---------+---------+---------+ 600
    AGGTTAGACGATTCAACCGACTGGGTCCAAGAGTTGGTGTGGTTCGAAGAGGTCAATAAC
                                CP-10
                                  CP-11
      V  I  I  P  E  G  S  G  Y  N  N  T  L  D  H  G  T  C  T  A
    ACGTTATTATTCCAGAAGGaTCcGGTTACAACAACACTTTGGACCACGGTACTTGTACTG
601 ---------+---------+---------+---------+---------+---------+ 660
```

```
TGCAATAATAAGGTCTTCCtAGgCCAATGTTGTTGTGAAACCTGGTGCCATGAACATGAC

  F  E  D  S  E  L  G  D  D  V  E  A  N  F  T  A  L  F  A  P

CTTTCGAAGACTCTGAATTGGGTGACGACGTTGAAGCTAACTTCACTGCTTTGTTCGCTC
661 ---------+---------+---------+---------+---------+---------+ 720
GAAAGCTTCTGAGACTTAACCCACTGCTGCAACTTCGATTGAAGTGACGAAACAAGCGAG
                                                      CP-12
  A  I  R  A  R  L  E  A  D  L  P  G  V  T  L  T  D  E  D  V

CAGCTATTAGAGCTAGATTGGAAGCTGACTTGCCAGGTGTTACTTTGACTGACGAAGACG
721 ---------+---------+---------+---------+---------+---------+ 780
GTCGATAATCTCGATCTAACCTTCGACTGAACGGTCCACAATGAAACTGACTGCTTCTGC

CP-13
  V  Y  L  M  D  M  C  P  F  E  T  V  A  R  T  S  D  A  T  E

TTGTTTACTTGATGGACATGTGTCCATTCGAAACTGTTGCTAGAACTTCTGACGCTACTG
781 ---------+---------+---------+---------+---------+---------+ 840
AACAAATGAACTACCTGTACACAGGTAAGCTTTGACAACGATCTTGAAGACTGCGATGAC

  L  S  P  F  C  A  L  F  T  H  D  E  W  R  Q  Y  D  Y  L  Q

AATTGTCTCCATTCTGTGCTTTGTTCACTCACGACGAATGGAGACAATACGACTACTTGC
841 ---------+---------+---------+---------+---------+---------+ 900
TTAACAGAGGTAAGACACGAAACAAGTGAGTGCTGCTTACCTCTGTTATGCTGATGAACG

    CP-14
      CP-15
  S  L  G  K  Y  Y  G  Y  G  A  G  N  P  L  G  P  A  Q  G  V

AATCTTTGGGTAAGTACTACGGTTACGGTGCTGGTAACCCATTGGGTCCAGCTCAAGGTG
901 ---------+---------+---------+---------+---------+---------+ 960
TTAGAAACCCATTCATGATGCCAATGCCACGACCATTGGGTAACCCAGGTCGAGTTCCAC

  G  F  A  N  E  L  I  A  R  L  T  R  S  P  V  Q  D  H  T  S

TTGGTTTCGCTAACGAATTGATTGCTAGATTGACTAGATCTCCAGTTCAAGACCACACTT
```

961 ---------+---------+---------+---------+---------+---------+
1020

AACCAAAGCGATTGCTTAACTAACGATCTAACTGATCTAGAGGTCAAGTTCTGGTGTGAA

CP-16
5 CP-17

T N H T L D S N P A T F P L N A T L Y A

CTACTAACCACACTTTGGACTCTAACCCAGCTACTTTCCCATTGAACGCTACTTTGTACG

1021 ---------+---------+---------+---------+---------+---------+
1080

10 GATGATTGGTGTGAAACCTGAGATTGGGTCGATGAAAGGGTAACTTGCGATGAAACATGC

D F S H D N S M I S I F F A L G L Y N G

CTGACTTCTCTCACGACAACTCTATGATTTCTATTTTCTTCGCTTTGGGTTTGTACAACG

1081 ---------+---------+---------+---------+---------+---------+
15 1140

GACTGAAGAGAGTGCTGTTGAGATACTAAAGATAAAAGAAGCGAAACCCAAACATGTTGC

CP-18
CP-19

T A P L S T T S V E S I E E T D G Y S A

20 GTACTGCTCCATTGTCTACTACTTCTGTTGAATCTATTGAAGAAACTGACGGTTACTCTG

1141 ---------+---------+---------+---------+---------+---------+
1200

CATGACGAGGTAACAGATGATGAAGACAACTTAGATAACTTCTTTGACTGCCAATGAGAC

25 S W T V P F G A R A Y V E M M Q C Q A E

CTTCTTGGACTGTTCCATTCGGTGCTAGAGCTTACGTTGAAATGATGCAATGTCAAGCTG

1201 ---------+---------+---------+---------+---------+---------+
1260

GAAGAACCTGACAAGGTAAGCCACGATCTCGAATGCAACTTTACTACGTTACAGTTCGAC

30 CP-20
CP-21

K E P L V R V L V N D R V V P L H G C A

AAAAGGAACCATTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTG

1261 ---------+---------+---------+---------+---------+---------+
35 1320

```
                TTTTCCTTGGTAACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACAC

                  V  D  K  L  G  R  C  K  R  D  D  F  V  E  G  L  S  F  A  R

                CTGTTGACAAGTTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTCGCTA

     1321 ---------+---------+---------+---------+---------+---------+
1380

                GACAACTGTTCAACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGAT

                                                                  CP-22
                  S  G  G  N  W  A  E  C  F  A  *  Eco RI

                GATCTGGTGGTAACTGGGCTGAATGTTTCGCTTAAGAATTCATATA

     1381 ---------+---------+---------+---------+------ 1426

                CTAGACCACCATTGACCCGACTTACAAAGCGAATTCTTAAGTATAT
```

Figure 3

```
                          1
50

P. involutus (phyA1) SvP.KnTAPt FPIPeseQrn WSPYSPYFPL AeYkAPPAGC
QInQVNIIQR

P. involutus (phyA2) SvP.RniAPK FSIPeseQrn WSPYSPYFPL AeYkAPPAGC
EInQVNIIQR

T. pubescens         hiPlRdTSAc LdVTrDvQqs WSmYSPYFPa AtYvAPPASC
QInQVHIIQR

A. pediades          GgvvQaTfvQ pfFPpQiQds WAAYTPYYPV qaYtPPPkDC
KItQVNIIQR

P. lycii             StQfsfvAAQ LPIPaQntsn WGPYdPFFPV EpYaAPPEGC
tVtQVNLIQR

Basidio              S-P-R-TAAQ LPIP-Q-Q-- WSPYSPYFPV A-Y-APPAGC QI-
QVNIIQR
```

```
                                51
    100
    P. involutus (phyA1) HGARFPTSGA TTRIKAGLTK LQGvqnfTDA KFNFIkSfkY
    dLGnsDLVPF
    P. involutus (phyA2) HGARFPTSGA ATRIKAGLSK LQSvqnfTDP KFDFIkSfTY
    dLGtsDLVPF
    T. pubescens         HGARFPTSGA AkRIQTAVAK LKAAsnyTDP lLAFVtNyTY
    sLGqDsLVeL
    A. pediades          HGARFPTSGA GTRIQAAVkK LQSAktyTDP RLDFLtNyTY
    tLGhDDLVPF
    P. lycii             HGARWPTSGA rSRqvAAVAK IQmArpfTDP KYEFLnDfvY
    kFGvADLLPF

    Basidio              HGARFPTSGA ATRIQAAVAK LQSA---TDP KLDFL-N-TY -LG-
    DDLVPF

                                101
    150
    P. involutus (phyA1) GAaQSfDAGQ EAFARYSkLV SkNNLPFIRA dGSDRVVDSA
    TNWTAGFAsA
    P. involutus (phyA2) GAaQSfDAGl EvFARYSkLV SsDNLPFIRS dGSDRVVDTA
    TNWTAGFAsA
    T. pubescens         GAtQSSEAGQ EAFTRYSsLV SaDELPFVRA SGSDRVVATA
    nNWTAGFAlA
    A. pediades          GAlQSSQAGE ETFqRYSfLV SkENLPFVRA SSSNRVVDSA
    TNWTEGFSaA
    P. lycii             GAnQShQTGt DmYTRYStLf egGDVPFVRA AGdQRVVDSS
    TNWTAGFGdA

    Basidio              GA-QSSQAGQ EAFTRYS-LV S-DNLPFVRA SGSDRVVDSA
    TNWTAGFA-A

                                151
    200
    P. involutus (phyA1) ShNTvqPkLn LILPQtGNDT LEDNMCPaAG DSDPQvNaWL
    AVafPSITAR
    P. involutus (phyA2) SrNAiqPkLd LILPQtGNDT LEDNMCPaAG ESDPQvDaWL
    AsafPSVTAQ
```

T. pubescens SsNSitPvLs VIISEaGNDT LDDNMCPaAG DSDPQvNqWL AqFAPPMTAR

A. pediades ShHvlnPiLf VILSEslNDT LDDaMCPnAG sSDPQtGiWt SIYGTPIAnR

P. lycii SgETvlPtLq VVLqEeGNcT LcNNMCPnEv DGDest.tWL GVFAPnITAR

Basidio **S-NT--P-L- VILSE-GNDT LDDNMCP-AG DSDPQ-N-WL AVFAPPITAR**

201

250

P. involutus (*phyA1*) LNAAAPSvNL TDtDAfNLvs LCAFlTVSkE kkSdFCtLFE giPGsFeAFa

P. involutus (*phyA2*) LNAAAPGANL TDaDAfNLvs LCPFmTVSkE qkSdFCtLFE giPGsFeAFa

T. pubescens LNAGAPGANL TDtDTyNLlt LCPFETVAtE rrSeFCDIYE elQAE.dAFa

A. pediades LNqqAPGANI TAaDvsNLip LCAFETIvkE tpSpFCNLF. .tPEEFaqFe

P. lycii LNAAAPSANL SDsDAltLmd MCPFDTLSsG naSpFCDLF. .tAEEYvSYe

Basidio **LNAAAPGANL TD-DA-NL-- LCPFETVS-E --S-FCDLFE --PEEF-AF-**

251

300

P. involutus (*phyA1*) YgGDLDKFYG TGYGQeLGPV QGVGYVNELI ARLTnsAVRD NTQTNRTLDA

P. involutus (*phyA2*) YaGDLDKFYG TGYGQALGPV QGVGYINELL ARLTnsAVnD NTQTNRTLDA

T. pubescens YnADLDKFYG TGYGQPLGPV QGVGYINELI ARLTaQnVsD HTQTNsTLDS

A. pediades YfGDLDKFYG TGYGQPLGPV QGVGYINELL ARLTemPVRD NTQTNRTLDS

P. lycii YyyDLDKYYG TGpGNALGPV QGVGYVNELL ARLTgQAVRD ETQTNRTLDS

Basidio **Y-GDLDKFYG TGYGQPLGPV QGVGYINELL ARLT-QAVRD**
NTQTNRTLDS

301
350

P. involutus (*phyA1*) SPvTFPLNKT FYADFSHDNl MVAVFSAMGL FrQPAPLsTS
vPNPwRTWrT

P. involutus (*phyA2*) APdTFPLNKT MYADFSHDNl MVAVFSAMGL FrQSAPLsTS
tPDPNRTWLT

T. pubescens SPeTFPLNRT LYADFSHDNQ MVAIFSAMGL FNQSAPLDPT
tPDPaRTFLv

A. pediades SPlTFPLDRS IYADLSHDNQ MIAIFSAMGL FNQSSPLDPS
fPNPKRTWVT

P. lycii dPaTFPLNRT FYADFSHDNt MVPIFAALGL FNaTA.LDPl
kPDeNRlWVd

Basidio **SP-TFPLNRT FYADFSHDNQ MVAIFSAMGL FNQSAPLDPS -**
PDPNRTWVT

351
400

P. involutus (*phyA1*) SsLVPFSGRM VVERLsC..f GT..... .....tkV
RVLVQDqVQP

P. involutus (*phyA2*) SsVVPFSARM aVERLsC..a GT..... .....tkV
RVLVQDqVQP

T. pubescens kKIVPFSARM VVERLdC..g GA..... .....qsV
RLLVNDAVQP

A. pediades SRLtPFSARM VtERLlCqrd GTgsggpsri mrngnvqtfV
RILVNDALQP

P. lycii SKLVPFSGHM tVEKLaC... ....sgkeaV
RVLVNDAVQP

Basidio **SKLVPFSARM VVERL-C--- GT-------- ---------V**
RVLVNDAVQP

401 441

P. involutus (*phyA1*)	LEFCGGDrNG lCTLAkFVES QtFARsDGaG DFEKCFATSa	~
P. involutus (*phyA2*)	LEFCGGDqDG lCALDkFVES QaYARsGGaG DFEKCLATTv	~
T. pubescens	LAFCGADtsG vCTLDAFVES QaYARNDGEG DFEKCFAT~~	~
A. pediades	LKFCGGDmDS lCTLEAFVES QkYAREDGQG DFEKCFD~~~	~
P. lycii	LEFCGG.vDG vCeLsAFVES QtYARENGQG DFAKCgfvPs	e
Basidio	**LEFCGGD-DG -CTLDAFVES Q-YAREDGQG DFEKCFATP-**	**-**

Modtaget PD

22 JAN. 1999

Figure 4

	1				50
A. terreus 9a1	KhsdCNSVDh	GYQCfPELSH	kWGlYAPYFS	LqDESPFPlD	VPeDCHITFV
A. terreus cbs	NhsdCtSVDr	GYQCfPELSH	kWGlYAPYFS	LqDESPFPlD	VPdDCHITFV
A. niger var. *awamori*	NqsTCDTVDq	GYQCfSEtSH	LWGQYAPFFS	LANESAISPD	VPaGCRVTFa
A. niger NRRL3135	NqsSCDTVDq	GYQCfSEtSH	LWGQYAPFFS	LANESvISPE	VPaGCRVTFa
A. fumigatus 13073	GSkSCDTVDl	GYQCsPAtSH	LWGQYSPFFS	LEDElSVSSK	LPkDCRITLV
A. fumigatus 32722	GSkSCDTVDl	GYQCsPAtSH	LWGQYSPFFS	LEDElSVSSK	LPkDCRITLV
A. fumigatus 58128	GSkSCDTVDl	GYQCsPAtSH	LWGQYSPFFS	LEDElSVSSK	LPkDCRITLV
A. fumigatus 26906	GSkSCDTVDl	GYQCsPAtSH	LWGQYSPFFS	LEDElSVSSK	LPkDCRITLV
A. fumigatus 32239	GSkACDTVEl	GYQCsPGtSH	LWGQYSPFFS	LEDElSVSSD	LPkDCRVTFV
E. nidulans	QNHSCNTaDG	GYQCfPNVSH	VWGQYSPYFS	IEQESAISeD	VPhGCeVTFV
T. thermophilus	DSHSCNTVEG	GYQCrPEISH	sWGQYSPFFS	LADQSEISPD	VPqNCKITFV
T. lanuginosa	----------	----nvDIAR	hWGQYSPFFS	LAEvSEISPA	VPkGCRVeFV
M. thermophila	ESRPCDTpDl	GFQCgTAISH	FWGQYSPYFS	VPsElDaS..	IPdDCeVTFa
Basidio	xSxPxrxtAA	qLPipxQxqx	xWSPYSPYFP	VAxyxA....	pPaGCQIxqV
Consensus	NSHSCDTVDG	GYQC-PEISH	LWGQYSPFFS	LADESAISPD	VP-GCRVTFV
Fcp10	**NSHSCDTVDG**	**GYQCFPEISH**	**LWGQYSPFFS**	**LADESAISPD**	**VPKGCRVTFV**

51

100

A. terreus 9a1	QVLARHGARs PThSKTKaYA AtIaAIQKSA TaFpGKYAFL QSYNYSLDSE	
A. terreus cbs	QVLARHGARs PTdSKTKaYA AtIaAIQKNA TaLpGKYAFL KSYNYSMGSE	
A. niger var. *awamori*	QVLSRHGARY PTeSKGKKYS ALIeEIQQNv TtFDGKYAFL KTYNYSLGAD	
A. niger NRRL3135	QVLSRHGARY PTdSKGKKYS ALIeEIQQNA TtFDGKYAFL KTYNYSLGAD	
A. fumigatus 13073	QVLSRHGARY PTSSKSKKYk kLVtAIQaNA TdFKGKFAFL KTYNYTLGAD	
A. fumigatus 32722	QVLSRHGARY PTSSKSKKYk kLVtAIQaNA TdFKGKFAFL KTYNYTLGAD	
A. fumigatus 58128	QVLSRHGARY PTSSKSKKYk kLVtAIQaNA TdFKGKFAFL KTYNYTLGAD	
A. fumigatus 26906	QVLSRHGARY PTSSKSKKYk kLVtAIQaNA TdFKGKFAFL KTYNYTLGAD	
A. fumigatus 32239	QVLSRHGARY PTASKSKKYk kLVtAIQKNA TeFKGKFAFL ETYNYTLGAD	
E. nidulans	QVLSRHGARY PTeSKSKaYS GLIeAIQKNA TsFwGQYAFL ESYNYTLGAD	
T. thermophilus	QLLSRHGARY PTSSKTElYS qLIsrIQKtA TaYKGyYAFL KdYrYqLGAN	
T. lanuginosa	QVLSRHGARY PTAhKSEvYA ELLqrIQDtA TeFKGDFAFL RdYaYhLGAD	
M. thermophila	QVLSRHGARa PTlkRAasYv DLIdrIHhGA isYgPgYEFL RTYDYTLGAD	
Basidio	NIIqRHGARF PTSGaAtRiq AaVakLQsax xxtDPKLDFL xnxtYxLGxD	
Consensus	QVLSRHGARY PTSSKSKKYS ALI-AIQKNA T-FKGKYAFL KTYNYTLGAD	
Fcp10	**QVLSRHGARY PTSSKSKKYS ALIEAIQKNA TAFKGKYAFL KTYNYTLGAD**	

	101				150
A. terreus 9a1	ELTPFGrNQL	rDlGaQFYeR	YNAL.TRhIn	PFVRATDAsR	VhESAEKFVE
A. terreus cbs	NLTPFGrNQL	qDlGaQFYRR	YDTL.TRhIn	PFVRAADSsR	VhESAEKFVE
A. niger var. *awamori*	DLTPFGEQEL	VNSGIKFYQR	YESL.TRnII	PFIRSSGSsR	VIASGEKFIE
A. niger NRRL3135	DLTPFGEQEL	VNSGIKFYQR	YESL.TRnIV	PFIRSSGSsR	VIASGKKFIE
A. fumigatus 13073	DLTPFGEQQL	VNSGIKFYQR	YKAL.ARsVV	PFIRASGSDR	VIASGEKFIE
A. fumigatus 32722	DLTPFGEQQL	VNSGIKFYQR	YKAL.ARsVV	PFIRASGSDR	VIASGEKFIE
A. fumigatus 58128	DLTPFGEQQL	VNSGIKFYQR	YKAL.ARsVV	PFIRASGSDR	VIASGEKFIE
A. fumigatus 26906	DLTAFGEQQL	VNSGIKFYQR	YKAL.ARsVV	PFIRASGSDR	VIASGEKFIE
A. fumigatus 32239	DLTPFGEQQM	VNSGIKFYQK	YKAL.AgsVV	PFIRSSGSDR	VIASGEKFIE
E. nidulans	DLTiFGENQM	VDSGaKFYRR	YKnL.ARknt	PFIRASGSDR	VVASAEKFIN
T. thermophilus	DLTPFGENQM	IQlGIKFYnH	YKSL.ARnaV	PFVRCSGSDR	VIASGrlFIE
T. lanuginosa	NLTRFGEEQM	MESGrQFYHR	YREq.AReIV	PFVRAAGSAR	VIASAEfFnr
M. thermophila	ELTRtGQQQM	VNSGIKFYRR	YRAL.ARksI	PFVRTAGqDR	VVhSAENFtQ
Basidio	DLvPFGAxQs	sQAGqEaFtR	YsxLvSxdnL	PFVRASGSDR	VVDSAtNWtA
Consensus	DLTPFGEQQM	VNSGIKFYRR	YKAL-AR-IV	PFVRASGSDR	VIASAEKFIE
Fcp10	**DLTPFGEQQM**	**VNSGIKFYRR**	**YKAL.ARKIV**	**PFVRASGSDR**	**VIASAEKFIE**

	151				200
A. terreus 9a1	GFQTARqDDh	hAnphQPSPr	VDVaIPEGsA	YNNTLEHSLC	TAFEs...St
A. terreus cbs	GFQNARqGDP	hAnphQPSPr	VDVVIPEGtA	YNNTLEHSIC	TAFEa...St
A. niger var. *awamori*	GFQSTKLkDP	rAqpgQSSPk	IDVVISEAsS	sNNTLDpGtC	TvFEd...SE
A. niger NRRL3135	GFQSTKLkDP	rAqpgQSSPk	IDVVISEAsS	sNNTLDpGtC	TvFEd...SE
A. fumigatus 13073	GFQqAKLADP	gAt.nRAAPa	ISVIIPESeT	FNNTLDHGVC	TkFEa...SQ
A. fumigatus 32722	GFQqAKLADP	gAt.nRAAPa	ISVIIPESeT	FNNTLDHGVC	TkFEa...SQ
A. fumigatus 58128	GFQqAKLADP	gAt.nRAAPa	ISVIIPESeT	FNNTLDHGVC	TkFEa...SQ
A. fumigatus 26906	GFQqAKLADP	gAt.nRAAPa	ISVIIPESeT	FNNTLDHGVC	TkFEa...SQ
A. fumigatus 32239	GFQqANVADP	gAt.nRAAPV	ISVIIPESeT	YNNTLDHSVC	TnFEa...SE
E. nidulans	GFRkAQLhDh	g.s.gQATPV	VNVIIPEidG	FNNTLDHStC	vSFEn...dE
T. thermophilus	GFQSAKVlDP	hSdkhDAPPt	INVIIeEGpS	YNNTLDtGsC	PvFEd...Ss
T. lanuginosa	GFQdAKdrDP	rSnkdQAePV	INVIISEEtG	sNNTLDgltC	PAaEe...Ap
M. thermophila	GFHSAlLADR	gStvrPTlPy	dmVVIPETaG	aNNTLHNDLC	TAFEegPySt
Basidio	GFaxA.....	..sxntxxPx	LxVILSExg.	.NDTLDDNMC	......PxAG
Consensus	GFQSAKLADP	-A---QASPV	INVIIPEG-G	YNNTLDHGLC	TAFE--P-SE
Fcp10	**GFQSAKLADP**	**GANPHQASPV**	**INVIIPEGAG**	**YNNTLDHGLC**	**TAFEE...SE**

201

250

A. terreus 9a1 VGDDavANFT AVFAPAIaqR LEAdLPGVQL StDDVVNLMA MCPFETVSlT

A. terreus cbs VGDAaADNFT AVFAPAIakR LEAdLPGVQL SADDVVNLMA MCPFETVSlT

A. niger var. *awamori* LADtVEANFT AtFAPSIRqR LEndLSGVtL TDtEVtyLMD MCSFDTIStS

A. niger NRRL3135 LADtVEANFT AtFvPSIRqR LEndLSGVtL TDtEVtyLMD MCSFDTIStS

A. fumigatus 13073 LGDEVAANFT ALFAPdIRAR aEkhLPGVtL TDEDVVSLMD MCSFDTVArT

A. fumigatus 32722 LGDEVAANFT ALFAPdIRAR aEkhLPGVtL TDEDVVSLMD MCSFDTVArT

A. fumigatus 58128 LGDEVAANFT ALFAPdIRAR aEkhLPGVtL TDEDVVSLMD MCSFDTVArT

A. fumigatus 26906 LGDEVAANFT ALFAPdIRAR aKkhLPGVtL TDEDVVSLMD MCSFDTVArT

A. fumigatus 32239 LGDEVEANFT ALFAPAIRAR IEkhLPGVQL TDDDVVSLMD MCSFDTVArT

E. nidulans rADEIEANFT AIMGPPIRkR LEndLPGIKL TNENVIyLMD MCSFDTMArT

T. thermophilus gGHDaQEKFA kqFAPAIlEK IKDhLPGVDL AvsDVpyLMD LCPFETLArn

T. lanuginosa .DptqpAEFl qVFGPRVlkK ItkhMPGVNL TlEDVplFMD LCPFDTVGsd

M. thermophila IGDDaQDtYl StFAGPItAR VNAnLPGaNL TDADtVaLMD LCPFETVAsS

Basidio dSDpqxnxWl AVFAPPItAR LNAaaPGaNL TDxDaxNLxx LCPFETVS..

Consensus LGDDVEANFT AVFAPPIRAR LEA-LPGVNL TDEDVVNLMD MCPFDTVA-T

Fcp10 LGDDVEANFT AVFAPPIRAR LEAHLPGVNL TDEDVVNLMD MCPFDTVART

251

300

A. terreus 9a1 dD..Aht... ......LSPF CDLFTa..tE WtQYNYLlSL dKYYGYGGGN

A. terreus cbs dD..Aht... ......LSPF CDLFTa..aE WtQYNYL1SL dKYYGYGGGN

A. niger var. awamori Tv..DTK... ......LSPF CDLFTH..dE WiHYDYLQSL kKYYGHGAGN

A. niger NRRL3135 Tv..DTK... ......LSPF CDLFTH..dE WiNYDYLQSL kKYYGHGAGN

A. fumigatus 13073 SD..ASQ... ......LSPF CQLFTH..nE WkKYNYLQSL gKYYGYGAGN

A. fumigatus 32722 SD..ASQ... ......LSPF CQLFTH..nE WkKYNYLQSL gKYYGYGAGN

A. fumigatus 58128 SD..ASQ... ......LSPF CQLFTH..nE WkKYNYLQSL gKYYGYGAGN

A. fumigatus 26906 SD..ASQ... ......LSPF CQLFTH..nE WkKYNYLQSL gKYYGYGAGN

A. fumigatus 32239 AD..ASE... ......LSPF CAIFTH..nE WkKYDYLQSL gKYYGYGAGN

E. nidulans AH..GTE... ......LSPF CAIFTE..kE WlQYDYLQSL sKYYGYGAGS

T. thermophilus ht..DT.... ......LSPF CALsTQ..eE WqaYDYYQSL gKYYGnGGGN

T. lanuginosa PvlfPrQ... ......LSPF CHLFTa..dD WmaYDYYyTL dKYYSHGGGS

M. thermophila SsdpATadag ggngrpLSPF CrLFSE..sE WraYDYLQSV gKWYGYGPGN

Basidio ....xexxSxF CDLFexxpeE FxaFxYxgdL dKFYGtGyGQ

Consensus SD--ATQ--- ------LSPF CDLFTH---E W-QYDYLQSL - KYYGYGAGN

Fcp10 SD..ATQ... ......LSPF CDLFTH..DE WIQYDYLQSL GKYYGYGAGN

301

350

A. terreus 9a1 PLGPvQGVGW aNELMARLTR A.PVHDHTCv NNTLDASPAT FPLNATLYAD

A. terreus cbs PLGPvQGVGW aNELIARLTR S.PVHDHTCv NNTLDANPAT FPLNATLYAD

A. niger var. *awamori* PLGPTQGVGY aNELIARLTH S.PVHDDTSS NHTLDSNPAT FPLNSTLYAD

A. niger NRRL3135 PLGPTQGVGY aNELIARLTH S.PVHDDTSS NHTLDSSPAT FPLNSTLYAD

A. fumigatus 13073 PLGPAQGIGF tNELIARLTR S.PVQDHTST NsTLvSNPAT FPLNATMYvD

A. fumigatus 32722 PLGPAQGIGF tNELIARLTR S.PVQDHTST NsTLvSNPAT FPLNATMYvD

A. fumigatus 58128 PLGPAQGIGF tNELIARLTR S.PVQDHTST NsTLvSNPAT FPLNATMYvD

A. fumigatus 26906 PLGPAQGIGF tNELIARLTR S.PVQDHTST NsTLvSNPAT FPLNATMYvD

A. fumigatus 32239 PLGPAQGIGF tNELIARLTN S.PVQDHTST NsTLDSDPAT FPLNATIYvD

E. nidulans PLGPAQGIGF tNELIARLTQ S.PVQDNTST NHTLDSNPAT FPLDrkLYAD

T. thermophilus PLGPAQGVGF vNELIARMTH S.PVQDYTTv NHTLDSNPAT FPLNATLYAD

T. lanuginosa AFGPSRGVGF vNELIARMTg NlPVKDHTTv NHTLDdNPET FPLDAvLYAD

M. thermophila PLGPTQGVGF vNELLARLA. GvPVRDgTST NRTLDGDPrT FPLGrPLYAD

Basidio PLGPvQGVGY iNELLARLTx qa.VRDNTqT NRTLDSSPxT FPLNrTFYAD

Consensus PLGPAQGVGF -NELIARLTH S-PVQDHTST NHTLDSNPAT FPLNATLYAD

Fcp10 PLGPAQGVGF VNELIARLTH S.PVQDHTST NHTLDSNPAT FPLNATLYAD

351

400

A. terreus 9a1 FSHDSnLVSI FWALGLYNGT aPLSqTSVE. .SvsQTDGYA AAWTVPFAAR

A. terreus cbs FSHDSnLVSI FWALGLYNGT kPLSqTTVE. .ditrTDGYA AAWTVPFAAR

A. niger var. *awamori* FSHDNGIISI LFALGLYNGT kPLSTTTVE. .NitQTDGFS SAWTVPFASR

A. niger NRRL3135 FSHDNGIISI LFALGLYNGT kPLSTTTVE. .NitQTDGFS SAWTVPFASR

A. fumigatus 13073 FSHDNSMVSI FFALGLYNGT ePLSrTSVE. .SaKElDGYS ASWvVPFGAR

A. fumigatus 32722 FSHDNSMVSI FFALGLYNGT gPLSrTSVE. .SaKElDGYS ASWvVPFGAR

A. fumigatus 58128 FSHDNSMVSI FFALGLYNGT ePLSrTSVE. .SaKElDGYS ASWvVPFGAR

A. fumigatus 26906 FSHDNSMVSI FFALGLYNGT ePLSrTSVE. .SaKElDGYS ASWvVPFGAR

A. fumigatus 32239 FSHDNGMIPI FFAMGLYNGT ePLSqTSeE. .StKESNGYS ASWAVPFGAR

E. nidulans FSHDNSMISI FFAMGLYNGT qPLSmdSVE. .SiQEmDGYA ASWTVPFGAR

T. thermophilus FSHDNTMtSI FaALGLYNGT akLSTTeIK. .SiEETDGYS AAWTVPFGGR

T. lanuginosa FSHDNTMtGI FsAMGLYNGT kPLSTSkIQP pTgAAADGYA ASWTVPFAAR

M. thermophila FSHDNdMMGV LgALGaYDGv pPLdkTA..R rdpEElGGYA ASWAVPFAAR

Basidio FSHDNqMVAI FsAMGLFNqS aPLdPSxpDP nrt.....Wv TSklVPFSAR

Consensus FSHDNTMVSI FFALGLYNGT -PLSTTSVEP -S-EETDGYA ASWTVPFAAR

Fcp10 FSHDNTMVSI FFALGLYNGT KPLSTTSVE. .SIEETDGYA ASWTVPFAAR

401

450

A. terreus 9a1 AYVEMMQC.. ra..... .....EKEPL VRVLVNDRVM PLHGCPtDKL

A. terreus cbs AYIEMMQC.. ra..... .....EKQPL VRVLVNDRVM PLHGCAVDNL

A. niger var. *awamori* lYVEMMQC.. Qa..... .....EQEPL VRVLVNDRVV PLHGCPIDaL

A. niger NRRL3135 lYVEMMQC.. Qa..... .....EQEPL VRVLVNDRVV PLHGCPVDaL

A. fumigatus 13073 AYfEtMQC.. Ks..... .....EKEPL VRaLINDRVV PLHGCDVDKL

A. fumigatus 32722 AYfEtMQC.. Ks..... .....EKEPL VRaLINDRVV PLHGCDVDKL

A. fumigatus 58128 AYfEtMQC.. Ks..... .....EKESL VRaLINDRVV PLHGCDVDKL

A. fumigatus 26906 AYfEtMQC.. Ks..... .....EKEPL VRaLINDRVV PLHGCDVDKL

A. fumigatus 32239 AYfEtMQC.. Ks..... .....EKEPL VRaLINDRVV PLHGCAVDKL

E. nidulans AYfELMQC.. E..... .....KKEPL VRVLVNDRVV PLHGCAVDKF

T. thermophilus AYIEMMQC.. Dd..... .....sDEPV VRVLVNDRVV PLHGCEVDsL

T. lanuginosa AYVELLRC.. Etetsseeee EG...EDEPF VRVLVNDRVV PLHGCrVDRW

M. thermophila iYVEkMRC.. sgggggggg EGrqeKDEeM VRVLVNDRVM TLkGCGaDEr

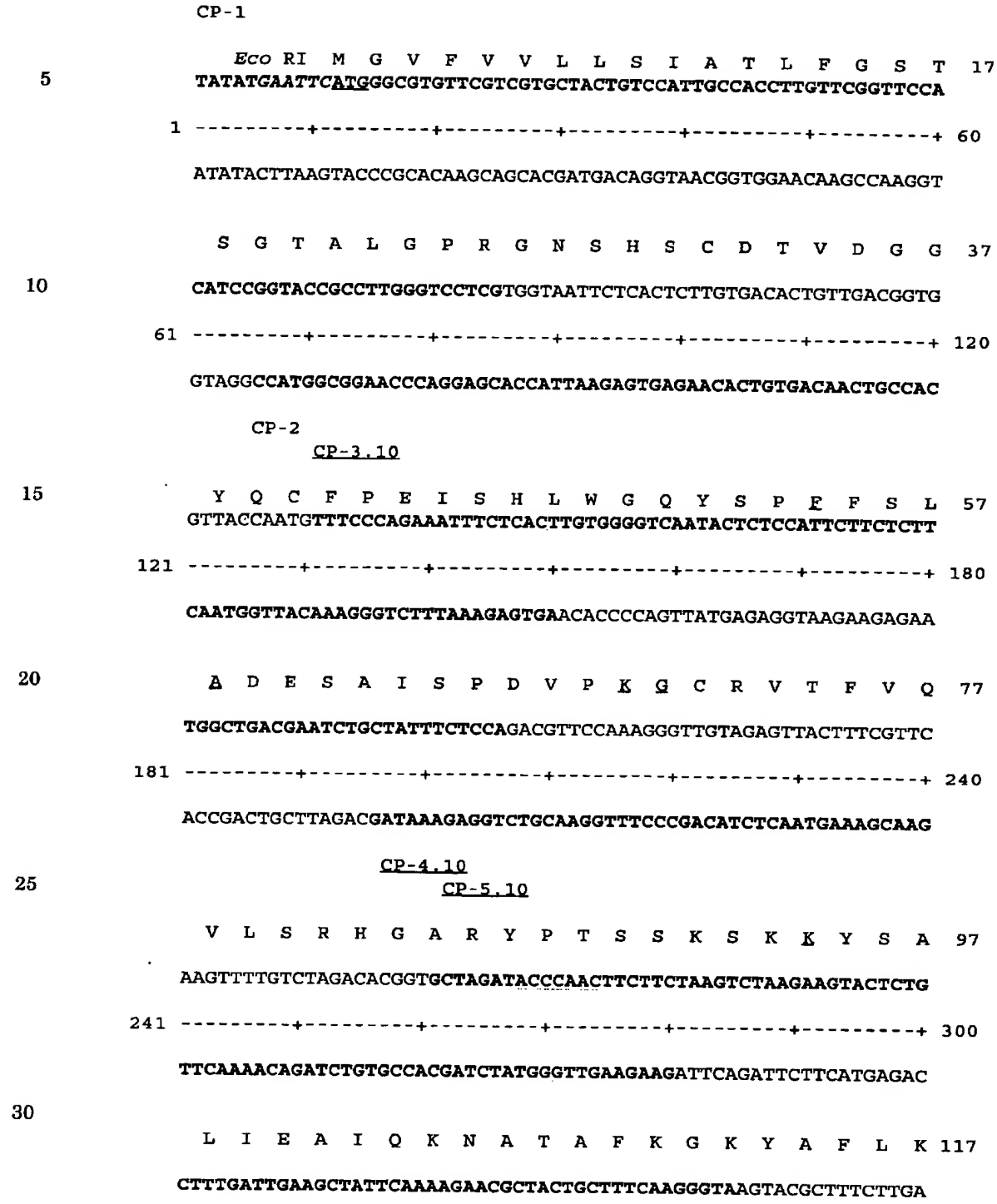
Basidio mvVErLxCxx xgtxxxxxxx xxxxxxxxxx VRVLVNDaVq PLEfCGgDxd

Consensus AYVEMMQC-- E--------- EG---EKEPL VRVLVNDRVV PLHGCGVDKL

Fcp10 AYVEMMQC.. EA..... .....EKEPL VRVLVNDRVV PLHGCGVDKL

```
                                    451                                 482
        A. terreus 9a1              GRCKrDAFVA GLSFAQAG.. GNWADCF~~~ ~~
        A. terreus cbs              GRCKrDDFVE GLSFARAG.. GNWAECF~~~ ~~
        A. niger var. awamori       GRCtrDsFVr GLSFARSG.. GDWAECsA~~ ~~
        A. niger NRRL3135           GRCtrDsFVr GLSFARSG.. GDWAECFA~~ ~~
        A. fumigatus 13073          GRCKlNDFVK GLSWARSG.. GNWGECFS~~ ~~
        A. fumigatus 32722          GRCKlNDFVK GLSWARSG.. GNWGECFS~~ ~~
        A. fumigatus 58128          GRCKlNDFVK GLSWARSG.. GNWGECFS~~ ~~
        A. fumigatus 26906          GRCKlNDFVK GLSWARSG.. GNWGECFS~~ ~~
        A. fumigatus 32239          GRCKlKDFVK GLSWARSG.. GNSEQSFS~~ ~~
        E. nidulans                 GRCTlDDWVE GLNFARSG.. GNWKtCFTl~ ~~
        T. thermophilus             GRCKrDDFVr GLSFARqG.. GNWEGCYAas e~
        T. lanuginosa               GRCRrDEWIK GLTFARqG.. GHWDrCF~~~ ~~
        M. thermophila              GmCTlErFIE SMAFARGN.. GKWDlCFA~~ ~~
        Basidio                     GxCTlDAFVE SqxYAReDgq GDFEKCFAtp xx

                        Consensus   GRCK-DDFVE GLSFARSG-- GNWEECFA-- --

                          Fcp10     GRCKRDDFVE GLSFARSG.. GNWEECFA.. ..
```

Figure 5



```
301 ---------+---------+---------+---------+---------+---------+ 360

    GAAACTAACTTCGATAAGTTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACT

                              CP-6
                                   CP-7.10

5             T  Y  N  Y  T  L  G  A  D  D  L  T  P  F  G  E  Q  Q  M  V 137

    AGACTTACAACTACACTTTGGGTGCTGACGACTTGACTCCATTCGGTGAACAACAAATGG

361 ---------+---------+---------+---------+---------+---------+ 420

    TCTGAATGTTGATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTGTTGTTTACC

10            N  S  G  I  K  F  Y  R  R  Y  K  A  L  A  R  K  I  V  P  F 157

    TTAACTCTGGTATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCAT

421 ---------+---------+---------+---------+---------+---------+ 480

    AATTGAGACCATAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTA

                                        CP-8.10

15                                           CP-9.10

              V  R  A  S  G  S  D  R  V  I  A  S  A  E  K  F  I  E  G  F 177
    TCGTTAGAGCTTCTGGTTCTGACAGAGTTATTGCTTCTGCTGAAAAGTTCATTGAAGGTT
481 ---------+---------+---------+---------+---------+---------+ 540
    AGCAATCTCGAAGACCAAGACTGTCTCAATAACGAAGACGACTTTTCAAGTAACTTCCAA

20

              Q  S  A  K  L  A  D  P  G  A  N  P  H  Q  A  S  P  V  I  N 197
    TCCAATCTGCTAAGTTGGCTGACCCAGGTGCTAACCCACACCAAGCTTCTCCAGTTATTA

541 ---------+---------+---------+---------+---------+---------+ 600

    AGGTTAGACGATTCAACCGACTGGGTCCACGATTGGGTGTGGTTCGAAGAGGTCAATAAT

25                                                  CP-10.10
                                                         CP-11.10

              V  I  I  P  E  G  A  G  Y  N  N  T  L  D  H  G  L  C  T  A 217
    ACGTTATTATTCCAGAAGGTGCTGGTTACAACAACACTTTGGACCACGGTTTGTGTACTG

601 ---------+---------+---------+---------+---------+---------+ 660

30  TGCAATAATAAGGTCTTCCACGACCAATGTTGTTGTGAAACCTGGTGCCAAACACATGAC

              F  E  E  S  E  L  G  D  D  V  E  A  N  F  T  A  V  F  A  P 237

    CTTTCGAAGAATCTGAATTGGGTGACGACGTTGAAGCTAACTTCACTGCTGTTTTCGCTC
```

```
661 ---------+---------+---------+---------+---------+---------+ 720
    GAAAGCTTCTTAGACTTAACCCACTGCTGCAACTTCGATTGAAGTGACGACAAAAGCGAG
                                                      CP-12.10
  5      P  I  R  A  R  L  E  A  H  L  P  G  V  N  L  T  D  E  D  V 257
    CACCTATTAGAGCTAGATTGGAAGCTCACTTGCCAGGTGTTAACTTGACTGACGAAGACG
721 ---------+---------+---------+---------+---------+---------+ 780
    GTGGATAATCTCGATCTAACCTTCGAGTGAACGGTCCACAATTGAACTGACTGCTTCTGC
 10      CP-13.10
       V  N  L  M  D  M  C  P  F  D  T  V  A  R  T  S  D  A  T  Q 277
    TTGTTAACTTGATGGACATGTGTCCATTCGACACTGTTGCTAGAACTTCTGACGCTACTC
781 ---------+---------+---------+---------+---------+---------+ 840
    AACAATTGAACTACCTGTACACAGGTAAGCTGTGACAACGATCTTGAAGACTGCGATGAG
 15
       L  S  P  F  C  D  L  F  T  H  D  E  W  I  Q  Y  D  Y  L  Q 297
    AATTGTCTCCATTCTGTGACTTGTTCACTCACGACGAATGGATTCAATACGACTACTTGC
841 ---------+---------+---------+---------+---------+---------+ 900
    TTAACAGAGGTAAGACACTGAACAAGTGAGTGCTGCTTACCTAAGTTATGCTGATGAACG
 20      CP-14.10
            CP-15.10
       S  L  G  K  Y  Y  G  Y  G  A  G  N  P  L  G  P  A  Q  G  V 317
    AATCTTTGGGTAAGTACTACGGTTACGGTGCTGGTAACCCATTGGGTCCAGCTCAAGGTG
901 ---------+---------+---------+---------+---------+---------+ 960
    TTAGAAACCCATTCATGATGCCAATGCCACGACCATTGGGTAACCCAGGTCGAGTTCCAC
 25
       G  F  V  N  E  L  I  A  R  L  T  H  S  P  V  Q  D  H  T  S 337
    TTGGTTTCGTTAACGAATTGATTGCTAGATTGACTCACTCTCCAGTTCAAGACCACACTT
961 ---------+---------+---------+---------+---------+---------+
 30 1020
    AACCAAAGCAATTGCTTAACTAACGATCTAACTGAGTGAGAGGTCAAGTTCTGGTGTGAA
            CP-16.10
               CP-17.10
```

```
                T  N  H  T  L  D  S  N  P  A  T  F  P  L  N  A  T  L  Y  A 357

                CTACTAACCACACTTTGGACTCTAACCCAGCTACTTTCCCATTGAACGCTACTTTGTACG

          1021 ---------+---------+---------+---------+---------+---------+
     1080

  5             GATGATTGGTGTGAAACCTGAGATTGGGTCGATGAAAGGGTAACTTGCGATGAAACATGC

                D  F  S  H  D  N  T  M  V  S  I  F  F  A  L  G  L  Y  N  G 377

                CTGACTTCTCTCACGACAACACTATGGTTTCTATTTTCTTCGCTTTGGGTTTGTACAACG

          1081 ---------+---------+---------+---------+---------+---------+
 10  1140

                GACTGAAGAGAGTGCTGTTGTGATACCAAAGATAAAAGAAGCGAAACCCAAACATGTTGC

                                          CP-18.10
                                               CP-19.10

                T  K  P  L  S  T  T  S  V  E  S  I  E  E  T  D  G  Y  A  A 397

 15             GTACTAAGCCATTGTCTACTACTTCTGTTGAATCTATTGAAGAAACTGACGGTTACGCTG

          1141 ---------+---------+---------+---------+---------+---------+
     1200

                CATGATTCGGTAACAGATGATGAAGACAACTTAGATAACTTCTTTGACTGCCAATGCGAC

 20             S  W  T  V  P  F  A  A  R  A  Y  V  E  M  M  Q  C  E  A  E 417

                CTTCTTGGACTGTTCCATTCGCTGCTAGAGCTTACGTTGAAATGATGCAATGTGAAGCTG
          1201 ---------+---------+---------+---------+---------+---------+
     1260

                GAAGAACCTGACAAGGTAAGCCACGATCTCGAATGCAACTTTACTACGTTACACTTCGAC

 25                                                  CP-20.10
                                                          CP-21.10

                K  E  P  L  V  R  V  L  V  N  D  R  V  V  P  L  H  G  C  G 437

                AAAAGGAACCATTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTG
          1261 ---------+---------+---------+---------+---------+---------+
 30  1320

                TTTTCCTTGGTAACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACAC


                V  D  K  L  G  R  C  K  R  D  D  F  V  E  G  L  S  F  A  R 457

                GTGTTGACAAGTTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTCGCTA
```

```
      1321 ---------+---------+---------+---------+---------+---------+
    1380

           CACAACTGTTCAACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGAT

                                                          CP-22.10
 5           S  G  G  N  W  E  E  C  F  A  *  Eco RI        467

           GATCTGGTGGTAACTGGGAAGAATGTTTCGCTTAAGAATTCATATA

      1381 ---------+---------+---------+---------+------ 1426

           CTAGACCACCATTGACCCTTCTTACAAAGCGAATTCTTAAGTATAT
```

10

Figure 6

```
                                1
    50
15  P. involutus (phyA1)        ~~~~~~~~~~ ~FPipeseqR nWSPYSPYFP LAEykA....
    pPaGCQInqV
    P. involutus (phyA2)        ~~~~~~~~~~ ~FsipeseqR nWSPYSPYFP LAEykA....
    pPaGCeInqV
    T. pubescens                ~~~~~~~~~~ ~LDvtRDVqQ sWSmYSPYFP aAtyvA....
20  pPaSCQInqV
    A. pediades                 ~~~~~~~~~~ ~pffpPQIqD sWAaYTPYYP VqAyTP....
    pPKDCKITqV
    P. lycii                    ~~~~~~~~~~ ~LPipAQnTs nWGPYdPFFP VEpyAA....
    pPEGCtVTqV
25  A. terreus 9a1              KhsdCNSVDh GYQCfPELSH kWGlYAPYFS LqDESPFPlD
    VPEDCHITFV
    A. terreus cbs              NhsdCtSVDr GYQCfPELSH kWGlYAPYFS LqDESPFPlD
    VPDDCHITFV
    A. niger var. awamori       NqsTCDTVDq GYQCfSEtSH LWGQYAPFFS LANESAISPD
30  VPaGCRVTFa
    A. niger T213               NqsSCDTVDq GYQCfSEtSH LWGQYAPFFS LANESvISPD
    VPaGCRVTFa
    A. niger NRRL3135           NqsSCDTVDq GYQCfSEtSH LWGQYAPFFS LANESvISPE
    VPaGCRVTFa
35  A. fumigatus ATCC13073      GSkSCDTVDl GYQCsPAtSH LWGQYSPFFS LEDElSVSSK
    LPKDCRITLV
```

A. fumigatus ATCC32722 GSkSCDTVDl GYQCsPAtSH LWGQYSPFFS LEDElSVSSK LPKDCRITLV

A. fumigatus ATCC58128 GSkSCDTVDl GYQCsPAtSH LWGQYSPFFS LEDElSVSSK LPKDCRITLV

A. fumigatus ATCC26906 GSkSCDTVDl GYQCsPAtSH LWGQYSPFFS LEDElSVSSK LPKDCRITLV

A. fumigatus ATCC32239 GSkACDTVEl GYQCsPGtSH LWGQYSPFFS LEDElSVSSD LPKDCRVTFV

E. nidulans QNHSCNTaDg GYQCfPNVSH VWGQYSPYFS IEQESAISeD VPhGCeVTFV

T. thermophilus DSHSCNTVEg GYQCrPEISH sWGQYSPFFS LADQSEISPD VPQNCKITFV

T. lanuginosa ---------- ----nvDIAR hWGQYSPFFS LAEvSEISPA VPKGCRVeFV

M. thermophila ESRPCDTpDl GFQCgTAISH FWGQYSPYFS VPsElDaS.. IPDDCeVTFa

Consensus Seq. 11 NSHSCDTVD- GYQC-PEISH LWGQYSPFFS LADESAISPD VPKGCRVTFV

51

100

P. involutus (*phyA1*) NIIqRHGARF PTSGaTtRik AgLtKLQgvq nftDAKFnFI KSFKYdLGns

P. involutus (*phyA2*) NIIqRHGARF PTSGaAtRik AgLsKLQsvq nftDPKFDFI KSFtYdLGTs

T. pubescens HIIqRHGARF PTSGaAKRiq TaVAKLKaaS nytDPlLAFV tnYtYSLGqD

A. pediades NIIqRHGARF PTSGaGtRiq AaVKKLQsak TytDPRLDFL tnYtYTLGhD

P. lycii NLIqRHGARW PTSGarsRqv AaVAKIQmar PftDPKYEFL NdFvYkFGvA

A. terreus 9a1 QVLARHGARs PThSKTKaYA AtIAaIQKSA TaFpGKYAFL QSYNYSLDSE

A. terreus cbs QVLARHGARs PTdSKTKaYA AtIAaIQKNA TaLpGKYAFL KSYNYSMGSE

A. niger var. *awamori* QVLSRHGARY PTeSKGKKYS ALIEeIQQNv TtFDGKYAFL KTYNYSLGAD

A. niger T213 QVLSRHGARY PTeSKGKKYS ALIEeIQQNv TtFDGKYAFL
KTYNYSLGAD

A. niger NRRL3135 QVLSRHGARY PTdSKGKKYS ALIEeIQQNA TtFDGKYAFL
KTYNYSLGAD

A. fumigatus ATCC13073 QVLSRHGARY PTSSKSKKYk kLVtaIQaNA TdFKGKFAFL
KTYNYTLGAD

A. fumigatus ATCC32722 QVLSRHGARY PTSSKSKKYk kLVtaIQaNA TdFKGKFAFL
KTYNYTLGAD

A. fumigatus ATCC58128 QVLSRHGARY PTSSKSKKYk kLVtaIQaNA TdFKGKFAFL
KTYNYTLGAD

A. fumigatus ATCC26906 QVLSRHGARY PTSSKSKKYk kLVtaIQaNA TdFKGKFAFL
KTYNYTLGAD

A. fumigatus ATCC32239 QVLSRHGARY PTASKSKKYk kLVtaIQKNA TeFKGKFAFL
ETYNYTLGAD

E. nidulans QVLSRHGARY PTeSKSKaYS GLIEaIQKNA TsFwGQYAFL
ESYNYTLGAD

T. thermophilus QLLSRHGARY PTSSKTElYS qLIsRIQKtA TaYKGyYAFL
KdYrYqLGAN

T. lanuginosa QVLSRHGARY PTAhKSEvYA ELLQRIQDtA TeFKGDFAFL
RdYaYhLGAD

M. thermophila QVLSRHGARa PTlkRAasYv DLIDRIHhGA isYgPgYEFL
RTYDYTLGAD

Consensus Seq. 11 QVLSRHGARY PTSSKSKKYS ALIERIQKNA T-FKGKYAFL
KTYNYTLGAD

101

150

P. involutus (*phyA1*) DLvPFGAaQs fDAGqEaFaR YskLvSKNnL PFIRAdGSDR
VVDSAtNWtA

P. involutus (*phyA2*) DLvPFGAaQs fDAGLEvFaR YskLvSsDnL PFIRSdGSDR
VVDTAtNWtA

T. pubescens sLveLGAtQs sEAGqEaFtR YsSLvSaDeL PFVRASGSDR
VVATANNWtA

A. pediades DLvPFGAlQs sQAGeEtFQR YsfLvSKEnL PFVRASSSNR
VVDSAtNWtE

P. lycii DLlPFGANQs hQTGtDMYtR YsTLfEgGdV PFVRAAGdQR
VVDSStNWtA

A. terreus 9a1 ELTPFGrNQL rDlGaQFYeR YNAL.TRHIn PFVRATDAsR VhESAEKFVE

A. terreus cbs NLTPFGrNQL qDlGaQFYRR YDTL.TRHIn PFVRAADSsR VhESAEKFVE

A. niger var. *awamori* DLTPFGEQEL VNSGIKFYQR YESL.TRNII PFIRSSGSsR VIASGEKFIE

A. niger T213 DLTPFGEQEL VNSGIKFYQR YESL.TRNII PFIRSSGSsR VIASGEKFIE

A. niger NRRL3135 DLTPFGEQEL VNSGIKFYQR YESL.TRNIV PFIRSSGSsR VIASGKKFIE

A. fumigatus ATCC13073 DLTPFGEQQL VNSGIKFYQR YKAL.ARSVV PFIRASGSDR VIASGEKFIE

A. fumigatus ATCC32722 DLTPFGEQQL VNSGIKFYQR YKAL.ARSVV PFIRASGSDR VIASGEKFIE

A. fumigatus ATCC58128 DLTPFGEQQL VNSGIKFYQR YKAL.ARSVV PFIRASGSDR VIASGEKFIE

A. fumigatus ATCC26906 DLTAFGEQQL VNSGIKFYQR YKAL.ARSVV PFIRASGSDR VIASGEKFIE

A. fumigatus ATCC32239 DLTPFGEQQM VNSGIKFYQK YKAL.AgSVV PFIRSSGSDR VIASGEKFIE

E. nidulans DLTiFGENQM VDSGaKFYRR YKnL.ARKnt PFIRASGSDR VVASAEKFIN

T. thermophilus DLTPFGENQM IQlGIKFYnH YKSL.ARNaV PFVRCSGSDR VIASGrlFIE

T. lanuginosa NLTRFGEEQM MESGrQFYHR YREq.AREIV PFVRAAGSAR VIASAEfFnr

M. thermophila ELTRtGQQQM VNSGIKFYRR YRAL.ARKsI PFVRTAGqDR VVhSAENFtQ

Consensus Seq. 11 **DLTPFGENQM VNSGIKFYRR YKAL-ARNIV PFVRASGSDR VIASAEKFIE**

151

200

P. involutus (*phyA1*) GFaSA..... ..shNtvqPk LNLILPQ..T gNDTLEDNMC PAaGD.....

P. involutus (*phyA2*) GFaSA..... ..srNaiqPk LDLILPQ..T gNDTLEDNMC PAaGE.....

T. pubescens PAaGD.....	GFalA..... ..ssNsiTPV	LSVIISE..A gNDTLDDNMC
A. pediades PnaGs.....	GFsAA..... ..shHvlNPI	LfVILSE..S LNDTLDDAMC
P. lycii PnevD.....	GFgdA..... ..sgEtvlPt	LQVVLQE..E gNcTLcNNMC
A. terreus 9a1 TAFEs...ST	GFQTARqDDh hAnpHQPSPr	VDVaIPEGSA YNNTLEHSLC
A. terreus cbs TAFEA...ST	GFQNARqGDP hAnpHQPSPr	VDVVIPEGTA YNNTLEHSIC
A. niger var. *awamori* TvFED...Se	GFQSTKLkDP rAqpgQSSPk	IDVVISEASS sNNTLDpGtC
A. niger T213 TvFED...Se	GFQSTKLkDP rAqpgQSSPk	IDVVISEASS sNNTLDpGtC
A. niger NRRL3135 TvFED...Se	GFQSTKLkDP rAqpgQSSPk	IDVVISEASS sNNTLDpGtC
A. fumigatus ATCC13073 TkFEA...Sq	GFQqAKLADP gAt.NRAAPa	ISVIIPESeT FNNTLDHGVC
A. fumigatus ATCC32722 TkFEA...Sq	GFQqAKLADP gAt.NRAAPa	ISVIIPESeT FNNTLDHGVC
A. fumigatus ATCC58128 TkFEA...Sq	GFQqAKLADP gAt.NRAAPa	ISVIIPESeT FNNTLDHGVC
A. fumigatus ATCC26906 TkFEA...Sq	GFQqAKLADP gAt.NRAAPa	ISVIIPESeT FNNTLDHGVC
A. fumigatus ATCC32239 TnFEA...Se	GFQqANVADP gAt.NRAAPV	ISVIIPESeT YNNTLDHSVC
E. nidulans vSFEN...de	GFRkAQLhDh g.s.gQATPV	VNVIIPEidG FNNTLDHStC
T. thermophilus PvFED...SS	GFQSAKVlDP hSdkHDAPPt	INVIIeEGPS YNNTLDtGsC
T. lanuginosa PAaEE...AP	GFQdAKdrDP rSnkDQAePV	INVIISEETG sNNTLDgltC
M. thermophila TAFEEgpyST	GFHSAlLADR gStvRPTlPy	dmVVIPETAG aNNTLHNDLC
Consensus Seq. 11 TAFED---ST	**GFQSAKLADP -A--HQASPV**	**INVIIPEGSG YNNTLDHGLC**

201

250

P. involutus (*phyA1*) .SDpqvnaWl AVafPSItAR LNAaaPSVNL TDtDafNLVs LCAFlTVSK.

P. involutus (*phyA2*) .SDpqvDaWl AsafPSVtAQ LNAaaPGaNL TDADafNLVs LCPFmTVSK.

T. pubescens .SDpqvnQWl AqFAPPMtAR LNAgaPGaNL TDtDtyNLLt LCPFETVAt.

A. pediades .SDpqtGiWT SIYGTPIanR LNqqaPGaNI TAADVsNLIp LCAFETIvK.

P. lycii .GDESt.tWl GVFAPnItAR LNAaaPSaNL SDsDaLtLMD MCPFDTLSs.

A. terreus 9a1 VGDDAvANFT AVFAPAIaqR LEAdLPGVQL StDDVVNLMA MCPFETVSlT

A. terreus cbs VGDAAADNFT AVFAPAIakR LEAdLPGVQL SADDVVNLMA MCPFETVSlT

A. niger var. *awamori* LADtvEANFT AtFAPSIRqR LEndLSGVtL TDtEVtyLMD MCSFDTIStS

A. niger T213 LADtvEANFT AtFAPSIRqR LEndLSGVtL TDtEVtyLMD MCSFDTIStS

A. niger NRRL3135 LADtvEANFT AtFvPSIRqR LEndLSGVtL TDtEVtyLMD MCSFDTIStS

A. fumigatus ATCC13073 LGDEvAANFT ALFAPdIRAR aEkhLPGVtL TDEDVVSLMD MCSFDTVART

A. fumigatus ATCC32722 LGDEvAANFT ALFAPdIRAR aEkhLPGVtL TDEDVVSLMD MCSFDTVART

A. fumigatus ATCC58128 LGDEvAANFT ALFAPdIRAR aEkhLPGVtL TDEDVVSLMD MCSFDTVART

A. fumigatus ATCC26906 LGDEvAANFT ALFAPdIRAR aKkhLPGVtL TDEDVVSLMD MCSFDTVART

A. fumigatus ATCC32239 LGDEvEANFT ALFAPAIRAR IEkhLPGVQL TDDDVVSLMD MCSFDTVART

E. nidulans rADEiEANFT AIMGPPIRkR LEndLPGIKL TNENVIyLMD MCSFDTMART

T. thermophilus gGHDAQEKFA kqFAPAIlEK IKDhLPGVDL AvsDVpyLMD LCPFETLARn

T. lanuginosa .DptqpAEFl qVFGPRVlkK ItkhMPGVNL TlEDVplFMD LCPFDTVGsd

M. thermophila IGDDAQDtYl StFAGPItAR VNAnLPGaNL TDADtVaLMD
LCPFETVAsS

Consensus Seq. 11 **LGDDAEANFT AVFAPPIRAR LEA-LPGVNL TDEDVVNLMD**
MCPFDTVART

251

300

P. involutus (*phyA1*) ....ekkSdF CtLFegiPGs FeaFAYggdL
dKFYGtGyGQ

P. involutus (*phyA2*) ....eqkSdF CtLFegiPGs FeaFAYagdL
dKFYGtGyGQ

T. pubescens ....errSeF CDIYeelqAE .daFAYnadL
dKFYGtGyGQ

A. pediades ....etpSPF CNLF..TPEE FaQFEYFgdL
dKFYGtGyGQ

P. lycii ....gnaSPF CDLF..TAEE YvsYEYYydL
dKYYGtGPGN

A. terreus 9a1 dD..Aht... ......LSPF CDLF..TAtE WtQYNYL1SL
dKYYGYGGGN

A. terreus cbs dD..Aht... ......LSPF CDLF..TAAE WtQYNYL1SL
dKYYGYGGGN

A. niger var. *awamori* Tv..DTK... ......LSPF CDLF..ThDE WiHYDYLQSL
kKYYGHGAGN

A. niger T213 Tv..DTK... ......LSPF CDLF..ThDE WiHYDYLRSL
kKYYGHGAGN

A. niger NRRL3135 Tv..DTK... ......LSPF CDLF..ThDE WiNYDYLQSL
kKYYGHGAGN

A. fumigatus ATCC13073 SD..ASQ... ......LSPF CQLF..ThNE WkKYNYLQSL
gKYYGYGAGN

A. fumigatus ATCC32722 SD..ASQ... ......LSPF CQLF..ThNE WkKYNYLQSL
gKYYGYGAGN

A. fumigatus ATCC58128 SD..ASQ... ......LSPF CQLF..ThNE WkKYNYLQSL
gKYYGYGAGN

A. fumigatus ATCC26906 SD..ASQ... ......LSPF CQLF..ThNE WkKYNYLQSL
gKYYGYGAGN

A. fumigatus ATCC32239 AD..ASE... ......LSPF CAIF..ThNE WkKYDYLQSL
gKYYGYGAGN

E. nidulans AH..GTE... ......LSPF CAIF..TEKE WlQYDYLQSL
sKYYGYGAGS

T. thermophilus ht..DT.... ......LSPF CALs..TqEE WqaYDYYQSL
gKYYGnGGGN

T. lanuginosa PvlfPrQ... ......LSPF CHLF..TADD WmaYDYYyTL
dKYYSHGGGS

M. thermophila SsdpATadag ggngrpLSPF CrLF..SEsE WraYDYLQSV
gKWYGYGPGN

Consensus Seq. 11 **SD--ATQ--- ------LSPF CDLF--TADE W-QYDYLQSL** -
KYYGYGAGN

301

350

P. involutus (*phyA1*) eLGPvQGVGY vNELIARLTN S.AVRDNTqT NRTLDASPvT
FPLNkTFYAD

P. involutus (*phyA2*) ALGPvQGVGY iNELLARLTN S.AVNDNTqT NRTLDAaPDT
FPLNkTMYAD

T. pubescens PLGPvQGVGY iNELIARLTa q.nVsDHTqT NsTLDSSPET
FPLNrTLYAD

A. pediades PLGPvQGVGY iNELLARLTE m.PVRDNTqT NRTLDSSPlT
FPLDrSIYAD

P. lycii ALGPvQGVGY vNELLARLTg q.AVRDETqT NRTLDSDPAT
FPLNrTFYAD

A. terreus 9a1 PLGPvQGVGW aNELMARLTR A.PVHDHTCv NNTLDASPAT
FPLNATLYAD

A. terreus cbs PLGPvQGVGW aNELIARLTR S.PVHDHTCv NNTLDANPAT
FPLNATLYAD

A. niger var. *awamori* PLGPTQGVGY aNELIARLTH S.PVHDDTSS NHTLDSNPAT
FPLNSTLYAD

A. niger T213 PLGPTQGVGY aNELIARLTH S.PVHDDTSS NHTLDSNPAT
FPLNSTLYAD

A. niger NRRL3135 PLGPTQGVGY aNELIARLTH S.PVHDDTSS NHTLDSSPAT
FPLNSTLYAD

A. fumigatus ATCC13073 PLGPAQGIGF tNELIARLTR S.PVQDHTST NsTLvSNPAT
FPLNATMYvD

A. fumigatus ATCC32722 PLGPAQGIGF tNELIARLTR S.PVQDHTST NsTLvSNPAT
FPLNATMYvD

A. fumigatus ATCC58128 PLGPAQGIGF tNELIARLTR S.PVQDHTST NsTLvSNPAT FPLNATMYvD

A. fumigatus ATCC26906 PLGPAQGIGF tNELIARLTR S.PVQDHTST NsTLvSNPAT FPLNATMYvD

A. fumigatus ATCC32239 PLGPAQGIGF tNELIARLTN S.PVQDHTST NsTLDSDPAT FPLNATIYvD

E. nidulans PLGPAQGIGF tNELIARLTQ S.PVQDNTST NHTLDSNPAT FPLDrkLYAD

T. thermophilus PLGPAQGVGF vNELIARMTH S.PVQDYTTv NHTLDSNPAT FPLNATLYAD

T. lanuginosa AFGPSRGVGF vNELIARMTg NlPVKDHTTv NHTLDdNPET FPLDAvLYAD

M. thermophila PLGPTQGVGF vNELLARLA. GvPVRDgTST NRTLDGDPrT FPLGrPLYAD

Consensus Seq. 11 PLGPAQGVGF -NELIARLTH S-PVQDHTST NHTLDSNPAT FPLNATLYAD

351

400

P. involutus (*phyA1*) FSHDNlMVAV FsAMGLFrqP aPLSTSvpNP wrt.....Wr TSSlVPFSGR

P. involutus (*phyA2*) FSHDNlMVAV FsAMGLFrqS aPLSTSTpDP nrt.....Wl TSSvVPFSAR

T. pubescens FSHDNqMVAI FsAMGLFNqS aPLdPTTpDP art.....Fl vkkiVPFSAR

A. pediades LSHDNqMIAI FsAMGLFNqS sPLdPSfpNP krt.....Wv TSRltPFSAR

P. lycii FSHDNTMVPI FaALGLFNAT a.LdPlkpDe nrl.....Wv DSklVPFSGH

A. terreus 9a1 FSHDSnLVSI FWALGLYNGT aPLSqTSVES Vs..QTDGYA AAWTVPFAAR

A. terreus cbs FSHDSnLVSI FWALGLYNGT KPLSqTTVEd It..rTDGYA AAWTVPFAAR

A. niger var. *awamori* FSHDNGIISI LFALGLYNGT KPLSTTTVEN It..QTDGFS SAWTVPFASR

A. niger T213 FSHDNGIISI LFALGLYNGT KPLSTTTVEN It..QTDGFS SAWTVPFASR

A. niger NRRL3135 FSHDNGIISI LFALGLYNGT KPLSTTTVEN It..QTDGFS
SAWTVPFASR

A. fumigatus ATCC13073 FSHDNSMVSI FFALGLYNGT EPLSrTSVES ak..ElDGYS
ASWvVPFGAR

A. fumigatus ATCC32722 FSHDNSMVSI FFALGLYNGT gPLSrTSVES ak..ElDGYS
ASWvVPFGAR

A. fumigatus ATCC58128 FSHDNSMVSI FFALGLYNGT EPLSrTSVES ak..ElDGYS
ASWvVPFGAR

A. fumigatus ATCC26906 FSHDNSMVSI FFALGLYNGT EPLSrTSVES ak..ElDGYS
ASWvVPFGAR

A. fumigatus ATCC32239 FSHDNGMIPI FFAMGLYNGT EPLSqTSeES tk..ESNGYS
ASWAVPFGAR

E. nidulans FSHDNSMISI FFAMGLYNGT QPLSmdSVES Iq..EmDGYA
ASWTVPFGAR

T. thermophilus FSHDNTMtSI FaALGLYNGT akLSTTeIKS Ie..ETDGYS
AAWTVPFGGR

T. lanuginosa FSHDNTMtGI FsAMGLYNGT KPLSTSkIQP ptgaAADGYA
ASWTVPFAAR

M. thermophila FSHDNdMMGV LgALGaYDGv pPLdkTArrd ..peElGGYA
ASWAVPFAAR

Consensus Seq. 11 **FSHDNTMVSI FFALGLYNGT KPLSTTSVES I---ETDGYA**
ASWTVPFAAR

401
450

P. involutus (*phyA1*) mvVErLsC.. fGt..... .....Tk VRVLVQDQVq
PLEfCGgDRn

P. involutus (*phyA2*) maVErLsC.. AGt..... .....Tk VRVLVQDQVq
PLEfCGgDQd

T. pubescens mvVErLDC.. GGa..... .....Qs VRLLVNDaVq
PLafCGaDts

A. pediades mvtErLlCQr DGtGsGGpsr imrNgnvQTF VRILVNDaLq
PLkfCGgDmd

P. lycii mtVEkLaC.. ....sgKea VRVLVNDaVq
PLEfCGg.vd

A. terreus 9a1 AYVEMMQCrA ....EK...EPL VRVLVNDRVM
PLHGCPtDKL

	A. terreus cbs PLHGCAVDNL	AYIEMMQCrA ....EK...QPL VRVLVNDRVM
	A. niger var. *awamori* PLHGCPIDaL	lYVEMMQCQA ....EQ...EPL VRVLVNDRVV
5	*A. niger* T213 PLHGCPIDaL	lYVEMMQCQA ....EQ...EPL VRVLVNDRVV
	A. niger NRRL3135 PLHGCPVDaL	lYVEMMQCQA ....EQ...EPL VRVLVNDRVV
	A. fumigatus ATCC13073	AYfEtMQCKS ....EK...EPL VRaLINDRVV
10	PLHGCDVDKL	
	A. fumigatus ATCC32722 PLHGCDVDKL	AYfEtMQCKS ....EK...EPL VRaLINDRVV
	A. fumigatus ATCC58128 PLHGCDVDKL	AYfEtMQCKS ....EK...ESL VRaLINDRVV
15	*A. fumigatus* ATCC26906 PLHGCDVDKL	AYfEtMQCKS ....EK...EPL VRaLINDRVV
	A. fumigatus ATCC32239 PLHGCAVDKL	AYfEtMQCKS ....EK...EPL VRaLINDRVV
	E. nidulans	AYfELMQCE. ....KK...EPL VRVLVNDRVV
20	PLHGCAVDKF	
	T. thermophilus PLHGCEVDsL	AYIEMMQCDD ....sD...EPV VRVLVNDRVV
	T. lanuginosa PLHGCrVDRW	AYVELLRCET ETsSeEEeEG ..ED...EPF VRVLVNDRVV
25	*M. thermophila* TLkGCGaDEr	iYVEkMRCsG GGgGgGGgEG ..rQekdEeM VRVLVNDRVM
	Consensus Seq. 11 PLHGCGVDKL	**AYVEMMQCEA GG-G-GG-EG --EK---EPL VRVLVNDRVV**

30

	451	482
P. involutus (*phyA1*)	GlCtLAKFVE SqTFARSDga GDFEKCFAts a~	
P. involutus (*phyA2*)	GlCaLDKFVE SqAYARSGga GDFEKCLAtt v~	
T. pubescens	GvCtLDAFVE SqAYARNDge GDFEKCFAt~ ~~	
A. pediades	SlCtLEAFVE SqkYAReDgq GDFEKCFD~~ ~~	
P. lycii	GvCELsAFVE SqTYAReNgq GDFAKCgfvp se	

A. terreus 9a1	GRCKrDAFVA	GLSFAQAG..	GNWADCF~~~	~~
A. terreus cbs	GRCKrDDFVE	GLSFARAG..	GNWAECF~~~	~~
A. niger var. *awamori*	GRCtrDsFVr	GLSFARSG..	GDWAECsA~~	~~
A. niger T213	GRCtrDsFVr	GLSFARSG..	GDWAECFA~~	~~
A. niger NRRL3135	GRCtrDsFVr	GLSFARSG..	GDWAECFA~~	~~
A. fumigatus ATCC13073	GRCKLNDFVK	GLSWARSG..	GNWGECFS~~	~~
A. fumigatus ATCC32722	GRCKLNDFVK	GLSWARSG..	GNWGECFS~~	~~
A. fumigatus ATCC58128	GRCKLNDFVK	GLSWARSG..	GNWGECFS~~	~~
A. fumigatus ATCC26906	GRCKLNDFVK	GLSWARSG..	GNWGECFS~~	~~
A. fumigatus ATCC32239	GRCKLKDFVK	GLSWARSG..	GNSEQSFS~~	~~
E. nidulans	GRCtLDDWVE	GLNFARSG..	GNWktCFTl~	~~
T. thermophilus	GRCKrDDFVr	GLSFARqG..	GNWEGCYAas	e~
T. lanuginosa	GRCRrDEWIK	GLTFARqG..	GHWDrCF~~~	~~
M. thermophila	GmCtLErFIE	SMAFARGN..	GKWDlCFA~~	~~
Consensus Seq. 11	**GRCKLDDFVE**	**GLSFARSG--**	**GNWAECFA--**	**--**

Figure 7

```
                M  G  V  F  V  V  L  L  S  I  A  T  L  F  G  S  T  S  G  T
        20
                ATGGGCGTGTTCGTCGTGCTACTGTCCATTGCCACCTTGTTCGGTTCCACATCCGGTACC
    5       1 ---+---------+---------+---------+---------+---------+---------+------
        60
                TACCCGCACAAGCAGCACGATGACAGGTAACGGTGGAACAAGCCAAGGTGTAGGCCATGG
                A  L  G  P  R  G  N  S  H  S  C  D  T  V  D  G  G  Y  Q  C
   10   40
                GCCTTGGGTCCTCGTGGTAATTCTCACTCTTGTGACACTGTTGACGGTGGTTACCAATGT
               61 ---+---------+---------+---------+---------+---------+---------+------
        120
                CGGAACCCAGGAGCACCATTAAGAGTGAGAACACTGTGACAACTGCCACCAATGGTTACA
   15
                F  P  E  I  S  H  L  W  G  T  Y  S  P  Y  F  S  L  A  D  E
        60
                TTCCCAGAAATTTCTCACTTGTGGGGTACCTACTCTCCATACTTCTCTTTGGCAGACGAA
              121 ---+---------+---------+---------+---------+---------+---------+------
   20   180
                AAGGGTCTTTAAAGAGTGAACACCCCATGGATGAGAGGTATGAAGAGAAACCGTCTGCTT
                S  A  I  S  P  D  V  P  D  D  C  R  V  T  F  V  Q  V  L  S
        80
   25           TCTGCTATTTCTCCAGACGTTCCAGACGACTGTAGAGTTACTTTCGTTCAAGTTTTGTCT
              187 ---+---------+---------+---------+---------+---------+---------+------
        240
                AGACGATAAAGAGGTCTGCAAGGTCTGCTGACATCTCAATGAAAGCAAGTTCAAAACAGA
   30           R  H  G  A  R  Y  P  T  S  S  A  S  K  A  Y  S  A  L  I  E
        100
                AGACACGGTGCTAGATACCCAACTTCTTCTGCGTCTAAGGCTTACTCTGCTTTGATTGAA
              241 ---+---------+---------+---------+---------+---------+---------+------
        300
   35           TCTGTGCCACGATCTATGGGTTGAAGAAGACGCAGATTCCGAATGAGACGAAACTAACTT
                A  I  Q  K  N  A  T  A  F  K  G  K  Y  A  F  L  K  T  Y  N
        120
```

```
                GCTATTCAAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGAAGACTTACAAC

          301 ---+---------+---------+---------+---------+---------+---------+------
    360

                CGATAAGTTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACTTCTGAATGTTG

 5

                Y  T  L  G  A  D  D  L  T  P  F  G  E  N  Q  M  V  N  S  G
    140

                TACACTTTGGGTGCTGACGACTTGACTCCATTCGGTGAAAACCAAATGGTTAACTCTGGT

          361 ---+---------+---------+---------+---------+---------+---------+------
10  420

                ATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTTTGGTTTACCAATTGAGACCA

                I  K  F  Y  R  R  Y  K  A  L  A  R  K  I  V  P  F  I  R  A
    160

15              ATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCATTCATTAGAGCT

          421 ---+---------+---------+---------+---------+---------+---------+------
    480

                TAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTAAGTAATCTCGA

20              S  G  S  D  R  V  I  A  S  A  E  K  F  I  E  G  F  Q  S  A
    180

                TCTGGTTCTGACAGAGTTATTGCTTCTGCTGAAAAGTTCATTGAAGGTTTCCAATCTGCT

          481 ---+---------+---------+---------+---------+---------+---------+------
    540

25              AGACCAAGACTGTCTCAATAACGAAGACGACTTTTCAAGTAACTTCCAAAGGTTAGACGA

                K  L  A  D  P  G  S  Q  P  H  Q  A  S  P  V  I  N  V  I  I
    200

                AAGTTGGCTGACCCAGGTTCTCAACCACACCAAGCTTCTCCAGTTATTAACGTGATCATT

30        541 ---+---------+---------+---------+---------+---------+---------+------
    600

                TTCAACCGACTGGGTCCAAGAGTTGGTGTGGTTCGAAGAGGTCAATAATTGCACTAGTAA

                P  E  G  S  G  Y  N  N  T  L  D  H  G  T  C  T  A  F  E  D
35  220

                CCAGAAGGATCCGGTTACAACAACACTTTGGACCACGGTACTTGTACTGCTTTCGAAGAC
```

```
      601 ---+---------+---------+---------+---------+---------+---------+------
      660

          GGTCTTCCTAGGCCAATGTTGTTGTGAAACCTGGTGCCATGAACATGACGAAAGCTTCTG

    5             S  E  L  G  D  D  V  E  A  N  F  T  A  L  F  A  P  A  I  R
      240

          TCTGAATTAGGTGACGACGTTGAAGCTAACTTCACTGCTTTGTTCGCTCCAGCTATTAGA
      661 ---+---------+---------+---------+---------+---------+---------+------
      720

   10         AGACTTAATCCACTGCTGCAACTTCGATTGAAGTGACGAAACAAGCGAGGTCGATAATCT

                A  R  L  E  A  D  L  P  G  V  T  L  T  D  E  D  V  V  Y  L
      260

          GCTAGATTGGAAGCTGACTTGCCAGGTGTTACTTTGACTGACGAAGACGTTGTTTACTTG
   15 721 ---+---------+---------+---------+---------+---------+---------+------
      780

          CGATCTAACCTTCGACTGAACGGTCCACAATGAAACTGACTGCTTCTGCAACAAATGAAC

                M  D  M  C  P  F  D  T  V  A  R  T  S  D  A  T  E  L  S  P
   20 280

          ATGGACATGTGTCCATTCGACACTGTCGCTAGAACTTCTGACGCTACTGAATTGTCTCCA
      781 ---+---------+---------+---------+---------+---------+---------+------
      840

          TACCTGTACACAGGTAAGCTGTGACAGCGATCTTGAAGACTGCGATGACTTAACAGAGGT
   25

                F  C  A  L  F  T  H  D  E  W  I  Q  Y  D  Y  L  Q  S  L  G
      300

          TTCTGTGCTTTGTTCACTCACGACGAATGGATCCAATACGACTACTTGCAAAGCTTGGGT
      841 ---+---------+---------+---------+---------+---------+---------+------
   30 900

          AAGACACGAAACAAGTGAGTGCTGCTTACCTAGGTTATGCTGATGAACGTTTCGAACCCA

                K  Y  Y  G  Y  G  A  G  N  P  L  G  P  A  Q  G  V  G  F  A
      320

   35         AAGTACTACGGTTACGGTGCTGGTAACCCATTGGGTCCAGCTCAAGGTGTTGGTTTCGCT
      901 ---+---------+---------+---------+---------+---------+---------+------
      960
```

```
                TTCATGATGCCAATGCCACGACCATTGGGTAACCCAGGTCGAGTTCCACAACCAAAGCGA

                N  E  L  I  A  R  L  T  H  S  P  V  Q  D  H  T  S  T  N  H
    340

 5              AACGAATTGATTGCTAGATTGACTCACTCTCCAGTTCAAGACCACACTTCTACTAACCAC

       961 ---+---------+---------+---------+---------+---------+---------+------
    1020

                TTGCTTAACTAACGATCTAACTGAGTGAGAGGTCAAGTTCTGGTGTGAAGATGATTGGTG

10              T  L  D  S  N  P  A  T  F  P  L  N  A  T  L  Y  A  D  F  S
    360

                ACTTTGGACTCTAACCCAGCTACTTTCCCATTGAACGCTACTTTGTACGCTGACTTCTCT

      1021 ---+---------+---------+---------+---------+---------+---------+------
    1080

15              TGAAACCTGAGATTGGGTCGATGAAAGGGTAACTTGCGATGAAACATGCGACTGAAGAGA

                H  D  N  T  M  I  S  I  F  F  A  L  G  L  Y  N  G  T  K  P
    380

                CACGACAACACTATGATATCTATTTTCTTCGCTTTGGGTTTGTACAACGGTACCAAGCCA

20    1081 ---+---------+---------+---------+---------+---------+---------+------
    1140

                GTGCTGTTGTGATACTATAGATAAAAGAAGCGAAACCCAAACATGTTGCCATGGTTCGGT

                L  S  T  T  S  V  E  S  I  E  E  T  D  G  Y  S  A  S  W  T
25  400

                TTGTCTACTACTTCTGTTGAATCTATTGAAGAAACTGACGGTTACTCTGCTTCTTGGACT

      1141 ---+---------+---------+---------+---------+---------+---------+------
    1200

                AACAGATGATGAAGACAACTTAGATAACTTCTTTGACTGCCAATGAGACGAAGAACCTGA

30
                V  P  F  A  A  R  A  Y  V  E  M  M  Q  C  Q  A  E  K  E  P
    420

                GTTCCATTCGCTGCTAGAGCTTACGTTGAAATGATGCAATGTCAAGCTGAAAAGGAACCA

      1201 ---+---------+---------+---------+---------+---------+---------+------
35  1260

                CAAGGTAAGCGACGATCTCGAATGCAACTTTACTACGTTACAGTTCGACTTTTCCTTGGT
```

```
                L  V  R  V  L  V  N  D  R  V  V  P  L  H  G  C  A  V  D  K
          440
                TTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTGCTGTTGACAAG
     1261 ---+---------+---------+---------+---------+---------+---------+------
          1320
                AACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACACGACAACTGTTC

                L  G  R  C  K  R  D  D  F  V  E  G  L  S  F  A  R  S  G  G
          460
                TTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTCGCTAGATCTGGTGGT
     1321 ---+---------+---------+---------+---------+---------+---------+------
          1380
                AACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGATCTAGACCACCA

                N  W  A  E  C  F  A  *    467
                AACTGGGCTGAATGTTTCGCTTAA
     1381 ---+---------+---------+ 1410
                TTGACCCGACTTACAAAGCGAATT
```

Figure 8

```
10
                M  G  V  F  V  V  L  L  S  I  A  T  L  F  G  S  T  S  G  T
20
                ATGGGCGTGTTCGTCGTGCTACTGTCCATTGCCACCTTGTTCGGTTCCACATCCGGTACC
              1 ---------+---------+---------+---------+---------+---------+
15  60
                TACCCGCACAAGCAGCACGATGACAGGTAACGGTGGAACAAGCCAAGGTGTAGGCCATGG
                A  L  G  P  R  G  N  S  H  S  C  D  T  V  D  G  G  Y  Q  C
    40
20              GCCTTGGGTCCTCGTGGTAACTCTCACTCTTGTGACACTGTTGACGGTGGTTACCAATGT
             61 ---------+---------+---------+---------+---------+---------+
    120
                CGGAACCCAGGAGCACCATTGAGAGTGAGAACACTGTGACAACTGCCACCAATGGTTACA
25  A           F  P  E  I  S  H  L  W  G  T  Y  S  P  F  F  S  L  A  D  E
    60
                TTCCCAGAAATTTCTCACTTGTGGGGTACATACTCTCCATTCTTCTCTTTGGCTGACGAA
            121 ---------+---------+---------+---------+---------+---------+
    180
30              AAGGGTCTTTAAAGAGTGAACACCCCATGTATGAGAGGTAAGAAGAGAAACCGACTGCTT
```

```
                S  A  I  S  P  D  V  P  K  G  C  R  V  T  F  V  Q  V  L  S
        80
                TCTGCTATTTCTCCAGACGTTCCAAAGGGTTGTAGAGTTACTTTCGTTCAAGTTTTGTCT
    5       181 ---------+---------+---------+---------+---------+---------+
        240
                AGACGATAAAGAGGTCTGCAAGGTTTCCCAACATCTCAATGAAAGCAAGTTCAAAACAGA
                R  H  G  A  R  Y  P  T  S  S  A  S  K  A  Y  S  A  L  I  E
   10   100
                AGACACGGTGCTAGATACCCAACTTCTTCTGCGTCTAAGGCGTACTCTGCTTTGATTGAA
            241 ---------+---------+---------+---------+---------+---------+
        300
                TCTGTGCCACGATCTATGGGTTGAAGAAGACGCAGATTCCGCATGAGACGAAACTAACTT
   15
                A  I  Q  K  N  A  T  A  F  K  G  K  Y  A  F  L  K  T  Y  N
        120
                GCTATTCAAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGAAGACTTACAAC
            301 ---------+---------+---------+---------+---------+---------+
   20   360
                CGATAAGTTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACTTCTGAATGTTG
        A       Y  T  L  G  A  D  D  L  T  P  F  G  E  Q  Q  M  V  N  S  G
        140
   25           TACACTTTGGGTGCTGACGACTTGACTCCATTCGGTGAACAACAAATGGTTAACTCTGGT
            361 ---------+---------+---------+---------+---------+---------+
        420
                ATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTGTTGTTTACCAATTGAGACCA
   30           I  K  F  Y  R  R  Y  K  A  L  A  R  K  I  V  P  F  I  R  A
        160
                ATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCATTCATTAGAGCT
            421 ---------+---------+---------+---------+---------+---------+
        480
   35           TAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTAAGTAATCTCGA
                S  G  S  D  R  V  I  A  S  A  E  K  F  I  E  G  F  Q  S  A
        180
```

```
                TCTGGTTCTGACAGAGTTATTGCTTCTGCTGAAAAGTTCATTGAAGGTTTCCAATCTGCT
          481 ---------+---------+---------+---------+---------+---------+
    540
                AGACCAAGACTGTCTCAATAACGAAGACGACTTTTCAAGTAACTTCCAAAGGTTAGACGA
 5
                K  L  A  D  P  G  A  N  P  H  Q  A  S  P  V  I  N  V  I  I
    200
                AAGTTGGCTGACCCAGGTGCTAACCCACACCAAGCTTCTCCAGTTATTAACGTTATTATT
          541 ---------+---------+---------+---------+---------+---------+
10  600
                TTCAACCGACTGGGTCCACGATTGGGTGTGGTTCGAAGAGGTCAATAATTGCAATAATAA
                P  E  G  A  G  Y  N  N  T  L  D  H  G  L  C  T  A  F  E  E
    220
15              CCAGAAGGTGCTGGTTACAACAACACTTTGGACCACGGTTTGTGTACTGCTTTCGAAGAA
          601 ---------+---------+---------+---------+---------+---------+
    660
                GGTCTTCCACGACCAATGTTGTTGTGAAACCTGGTGCCAAACACATGACGAAAGCTTCTT
20              S  E  L  G  D  D  V  E  A  N  F  T  A  V  F  A  P  P  I  R
    240
                TCTGAATTGGGTGACGACGTTGAAGCTAACTTCACTGCTGTTTTCGCTCCACCAATTAGA
          661 ---------+---------+---------+---------+---------+---------+
    720
25              AGACTTAACCCACTGCTGCAACTTCGATTGAAGTGACGACAAAAGCGAGGTGGTTAATCT
                A  R  L  E  A  H  L  P  G  V  N  L  T  D  E  D  V  V  N  L
    260
                GCTAGATTGGAAGCTCACTTGCCAGGTGTTAACTTGACTGACGAAGACGTTGTTAACTTG
30        721 ---------+---------+---------+---------+---------+---------+
    780
                CGATCTAACCTTCGAGTGAACGGTCCACAATTGAACTGACTGCTTCTGCAACAATTGAAC
                M  D  M  C  P  F  D  T  V  A  R  T  S  D  A  T  Q  L  S  P
35  280
                ATGGACATGTGTCCATTCGACACTGTTGCTAGAACTTCTGACGCTACTCAATTGTCTCCA
```

```
          781 ---------+---------+---------+---------+---------+---------+
     840

              TACCTGTACACAGGTAAGCTGTGACAACGATCTTGAAGACTGCGATGAGTTAACAGAGGT

  5           F  C  D  L  F  T  H  D  E  W  I  Q  Y  D  Y  L  Q  S  L  G
     300

              TTCTGTGACTTGTTCACTCACGACGAATGGATTCAATACGACTACTTGCAATCTTTGGGT

          841 ---------+---------+---------+---------+---------+---------+
     900

 10           AAGACACTGAACAAGTGAGTGCTGCTTACCTAAGTTATGCTGATGAACGTTAGAAACCCA

              K  Y  Y  G  Y  G  A  G  N  P  L  G  P  A  Q  G  V  G  F  V
     320

              AAGTACTACGGTTACGGTGCTGGTAACCCATTGGGTCCAGCTCAAGGTGTTGGTTTCGTT

 15       901 ---------+---------+---------+---------+---------+---------+
     960

              TTCATGATGCCAATGCCACGACCATTGGGTAACCCAGGTCGAGTTCCACAACCAAAGCAA

              N  E  L  I  A  R  L  T  H  S  P  V  Q  D  H  T  S  T  N  H
 20  340

              AACGAATTGATTGCTAGATTGACTCACTCTCCAGTTCAAGACCACACTTCTACTAACCAC

          961 ---------+---------+---------+---------+---------+---------+
    1020

              TTGCTTAACTAACGATCTAACTGAGTGAGAGGTCAAGTTCTGGTGTGAAGATGATTGGTG

 25

              T  L  D  S  N  P  A  T  F  P  L  N  A  T  L  Y  A  D  F  S
     360

              ACTTTGGACTCTAACCCAGCTACTTTCCCATTGAACGCTACTTTGTACGCTGACTTCTCT

         1021 ---------+---------+---------+---------+---------+---------+
 30  1080

              TGAAACCTGAGATTGGGTCGATGAAAGGGTAACTTGCGATGAAACATGCGACTGAAGAGA

              H  D  N  T  M  V  S  I  F  F  A  L  G  L  Y  N  G  T  K  P
     380

 35           CACGACAACACTATGGTTTCTATTTTCTTCGCTTTGGGTTTGTACAACGGTACTAAGCCA

         1081 ---------+---------+---------+---------+---------+---------+
    1140
```

```
                GTGCTGTTGTGATACCAAAGATAAAAGAAGCGAAACCCAAACATGTTGCCATGATTCGGT

                L  S  T  T  S  V  E  S  I  E  E  T  D  G  Y  S  A  S  W  T
    400

5               TTGTCTACTACTTCTGTTGAATCTATTGAAGAAACTGACGGTTACTCTGCTTCTTGGACT

        1141 ---------+---------+---------+---------+---------+---------+
    1200

                AACAGATGATGAAGACAACTTAGATAACTTCTTTGACTGCCAATGAGACGAAGAACCTGA

10              V  P  F  A  A  R  A  Y  V  E  M  M  Q  C  E  A  E  K  E  P
    420

                GTTCCATTCGCTGCTAGAGCTTACGTTGAAATGATGCAATGTGAAGCTGAAAAGGAACCA

        1201 ---------+---------+---------+---------+---------+---------+
    1260

15              CAAGGTAAGCGACGATCTCGAATGCAACTTTACTACGTTACACTTCGACTTTTCCTTGGT

                L  V  R  V  L  V  N  D  R  V  V  P  L  H  G  C  G  V  D  K
    440

                TTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTGGTGTTGACAAG

20      1261 ---------+---------+---------+---------+---------+---------+
    1320

                AACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACACCACAACTGTTC

                L  G  R  C  K  R  D  D  F  V  E  G  L  S  F  A  R  S  G  G
25  460

                TTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTCGCTAGATCTGGTGGT

        1321 ---------+---------+---------+---------+---------+---------+
    1380

                AACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGATCTAGACCACCA

30

                N  W  E  E  C  F  A  *     467

                AACTGGGAAGAATGTTTCGCTTAA

        1381 ---------+---------+---- 1404

                TTGACCCTTCTTACAAAGCGAATT

35
```

Figure 9

```
          M  G  V  F  V  V  L  L  S  I  A  T  L  F  G  S  T  S  G  T   20
          ATGGGGGTTTTCGTCGTTCTATTATCTATCGCGACTCTGTTCGGCAGCACATCGGGCACT
      1 ---------+---------+---------+---------+---------+---------+   60
          TACCCCCAAAAGCAGCAAGATAATAGATAGCGCTGAGACAAGCCGTCGTGTAGCCCGTGA
```

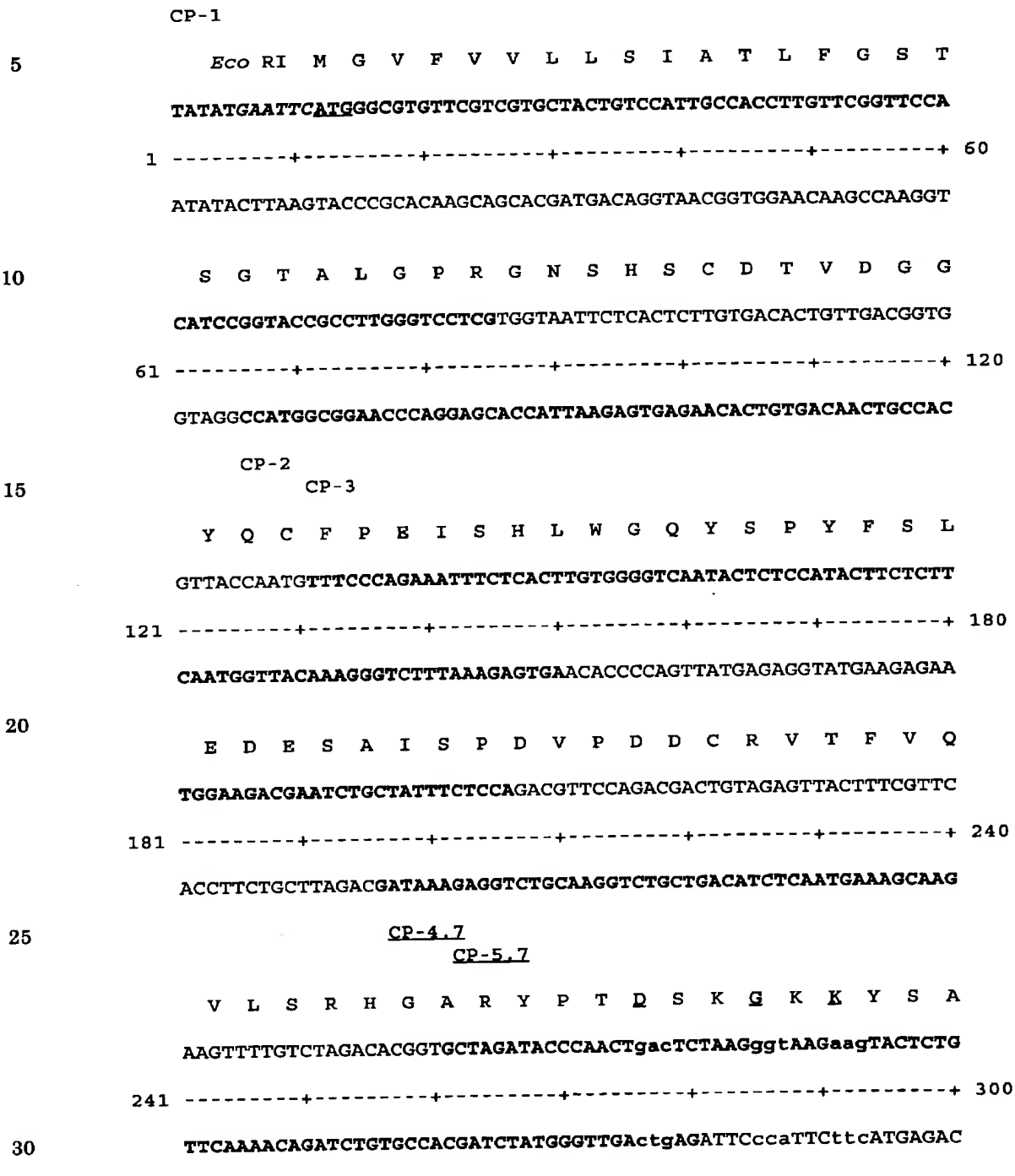
```
          A  L  G  P  R  G  N  H  S  K  S  C  D  T  V  D  L  G  Y  Q   40
          GCGCTGGGCCCCCGTGGAAATCACTCCAAGTCCTGCGATACGGTAGACCTAGGGTACCAG
      61 ---------+---------+---------+---------+---------+---------+ 120
          CGCGACCCGGGGGCACCTTTAGTGAGGTTCAGGACGCTATGCCATCTGGATCCCATGGTC
          C  S  P  A  T  S  H  L  W  G  T  Y  S  P  Y  F  S  L  E  D   60
          TGCTCCCCTGCGACTTCTCATCTATGGGGCACGTACTCGCCATaCTTTTCGCTCGAGGAC
     121 ---------+---------+---------+---------+---------+---------+ 180
          ACGAGGGGACGCTGAAGAGTAGATACCCCGtgCATGAGCGGTAtGAAAAGCGAGCTCCTG
          E  L  S  V  S  S  K  L  P  K  D  C  R  I  T  L  V  Q  V  L   80
          GAGCTGTCCGTGTCGAGTAAGCTTCCCAAGGATTGCCGGATCACCTTGGTACAGGTGCTA
     181 ---------+---------+---------+---------+---------+---------+ 240
          CTCGACAGGCACAGCTCATTCGAAGGGTTCCTAACGGCCTAGTGGAACCATGTCCACGAT
          S  R  H  G  A  R  Y  P  T  S  S  K  S  K  K  Y  K  K  L  I  100
          TCGCGCCATGGAGCGCGGTACCCAACCAGCTCCAAGAGCAAAAAGTATAAGAAGCTTaTt
     241 ---------+---------+---------+---------+---------+---------+ 300
          AGCGCGGTACCTCGCGCCATGGGTTGGTCGAGGTTCTCGTTTTTCATATTCTTCGAAtAa
          T  A  I  Q  A  N  A  T  D  F  K  G  K  Y  A  F  L  K  T  Y  120
          ACGGCGATCCAGGCCAATGCCACCGACTTCAAGGGCAAGTacGCCTTTTTGAAGACGTAC
     301 ---------+---------+---------+---------+---------+---------+ 360
          TGCCGCTAGGTCCGGTTACGGTGGCTGAAGTTCCCGTTCAtgCGGAAAAACTTCTGCATG
          N  Y  T  L  G  A  D  D  L  T  P  F  G  E  Q  Q  L  V  N  S  140
          AACTATACTCTGGGTGCGGATGACCTCACTCCCTTTGGGGAGCAGCAGCTGGTGAACTCG
     361 ---------+---------+---------+---------+---------+---------+ 420
          TTGATATGAGACCCACGCCTACTGGAGTGAGGGAAACCCCTCGTCGTCGACCACTTGAGC
          G  I  K  F  Y  Q  R  Y  K  A  L  A  R  S  V  V  P  F  I  R  160
          GGCATCAAGTTCTACCAGAGGTACAAGGCTCTGGCGCGCAGTGTGGTGCCGTTTATTCGC
```

```
421 ---------+---------+---------+---------+---------+---------+ 480
    CCGTAGTTCAAGATGGTCTCCATGTTCCGAGACCGCGCGTCACACCACGGCAAATAAGCG

        A  S  G  S  D  R  V  I  A  S  G  E  K  F  I  E  G  F  Q  Q   180
    GCCTCAGGCTCGGACCGGGTTATTGCTTCGGGAGAGAAGTTCATCGAGGGGTTCCAGCAG
481 ---------+---------+---------+---------+---------+---------+ 540
    CGGAGTCCGAGCCTGGCCCAATAACGAAGCCCTCTCTTCAAGTAGCTCCCCAAGGTCGTC

        A  K  L  A  D  P  G  A  T  N  R  A  A  P  A  I  S  V  I  I   200
    GCGAAGCTGGCTGATCCTGGCGCGACGAACCGCGCCGCTCCGGCGATTAGTGTGATTATT
541 ---------+---------+---------+---------+---------+---------+ 600
    CGCTTCGACCGACTAGGACCGCGCTGCTTGGCGCGGCGAGGCCGCTAATCACACTAATAA

        P  E  S  E  T  F  N  N  T  L  D  H  G  V  C  T  K  F  E  A   220
    CCGGAGAGCGAGACGTTCAACAATACGCTGGACCACGGTGTGTGCACGAAGTTTGAGGCG
601 ---------+---------+---------+---------+---------+---------+ 660
    GGCCTCTCGCTCTGCAAGTTGTTATGCGACCTGGTGCCACACACGTGCTTCAAACTCCGC

        S  Q  L  G  D  E  V  A  A  N  F  T  A  L  F  A  P  D  I  R   240
    AGTCAGCTGGGAGATGAGGTTGCGGCCAATTTCACTGCGCTCTTTGCACCCGACATCCGA
661 ---------+---------+---------+---------+---------+---------+ 720
    TCAGTCGACCCTCTACTCCAACGCCGGTTAAAGTGACGCGAGAAACGTGGGCTGTAGGCT

        A  R  L  E  K  H  L  P  G  V  T  L  T  D  E  D  V  V  S  L   260
    GCTCGCctCGAGAAGCATCTTCCTGGCGTGACGCTGACAGACGAGGACGTTGTCAGTCTA
721 ---------+---------+---------+---------+---------+---------+ 780
    CGAGCGgaGCTCTTCGTAGAAGGACCGCACTGCGACTGTCTGCTCCTGCAACAGTCAGAT

        M  D  M  C  P  F  D  T  V  A  R  T  S  D  A  S  Q  L  S  P   280
    ATGGACATGTGTcCGTTTGATACGGTAGCGCGCACCAGCGACGCAAGTCAGCTGTCACCG
781 ---------+---------+---------+---------+---------+---------+ 840
```

```
                TACCTGTACACAgGCAAACTATGCCATCGCGCGTGGTCGCTGCGTTCAGTCGACAGTGGC

                F  C  Q  L  F  T  H  N  E  W  K  K  Y  D  Y  L  Q  S  L  G   300

                TTCTGTCAACTCTTCACTCACAATGAGTGGAAGAAGTACgACTACCTTCAGTCCTTGGGC

5          841 ---------+---------+---------+---------+---------+---------+ 900

                AAGACAGTTGAGAAGTGAGTGTTACTCACCTTCTTCATGcTGATGGAAGTCAGGAACCCG

                K  Y  Y  G  Y  G  A  G  N  P  L  G  P  A  Q  G  I  G  F  T   320

                AAGTACTACGGCTACGGCGCAGGCAACCCTCTGGGACCGGCTCAGGGGATAGGGTTCACC

10         901 ---------+---------+---------+---------+---------+---------+ 960

                TTCATGATGCCGATGCCGCGTCCGTTGGGAGACCCTGGCCGAGTCCCCTATCCCAAGTGG

                N  E  L  I  A  R  L  T  R  S  P  V  Q  D  H  T  S  T  N  S
           340

15              AACGAGCTGATTGCCCGGTTGACgCGTTCGCCAGTGCAGGACCACACCAGCACTAACTCG

           961 ---------+---------+---------+---------+---------+---------+
      1020

                TTGCTCGACTAACGGGCCAACTGcGCAAGCGGTCACGTCCTGGTGTGGTCGTGATTGAGC

20              T  L  V  S  N  P  A  T  F  P  L  N  A  T  M  Y  V  D  F  S
           360

                ACTCTAGTCTCCAACCCGGCCACCTTCCCGTTGAACGCTACCATGTACGTCGACTTTTCA

          1021 ---------+---------+---------+---------+---------+---------+
      1080

25              TGAGATCAGAGGTTGGGCCGGTGGAAGGGCAACTTGCGATGGTACATGCAGCTGAAAAGT

                H  D  N  S  M  V  S  I  F  F  A  L  G  L  Y  N  G  T  E  P
           380

                CACGACAACAGCATGGTTTCCATCTTCTTTGCATTGGGCCTGTACAACGGCACTGAACCC

30        1081 ---------+---------+---------+---------+---------+---------+
      1140

                GTGCTGTTGTCGTACCAAAGGTAGAAGAAACGTAACCCGGACATGTTGCCGTGACTTGGG

                L  S  R  T  S  V  E  S  A  K  E  L  D  G  Y  S  A  S  W  V
35         400
```

```
                TTGTCCCGGACCTCGGTGGAAAGCGCCAAGGAATTGGATGGGTATTCTGCATCCTGGGTG

          1141 ---------+---------+---------+---------+---------+---------+
1200

                AACAGGGCCTGGAGCCACCTTTCGCGGTTCCTTAACCTACCCATAAGACGTAGGACCCAC

5

                V  P  F  G  A  R  A  Y  F  E  T  M  Q  C  K  S  E  K  E  P
420

                GTGCCTTTCGGCGCGCGAGCCTACTTCGAGACGATGCAATGCAAGTCGGAAAAGGAGCCT

          1201 ---------+---------+---------+---------+---------+---------+
10  1260

                CACGGAAAGCCGCGCGCTCGGATGAAGCTCTGCTACGTTACGTTCAGCCTTTTCCTCGGA

                L  V  R  A  L  I  N  D  R  V  V  P  L  H  G  C  D  V  D  K
440

15              CTTGTTCGCGCTTTGATTAATGACCGGGTTGTGCCACTGCATGGCTGCGATGTGGACAAG

          1261 ---------+---------+---------+---------+---------+---------+
1320

                GAACAAGCGCGAAACTAATTACTGGCCCAACACGGTGACGTACCGACGCTACACCTGTTC

20              L  G  R  C  K  L  N  D  F  V  K  G  L  S  W  A  R  S  G  G
460

                CTGGGGCGATGCAAGCTGAATGACTTTGTCAAGGGATTGAGTTGGGCCAGATCTGGGGGC

          1321 ---------+---------+---------+---------+---------+---------+
1380

25              GACCCCGCTACGTTCGACTTACTGAAACAGTTCCCTAACTCAACCCGGTCTAGACCCCCG


                N  W  G  E  C  F  S  *    467

                AACTGGGGAGAGTGCTTTAGTTGA

          1381 ---------+---------+---- 1404

30              TTGACCCCTCTCACGAAATCAACT
```

Figure 10



```
                L  I  E  A  I  Q  K  N  A  T  A  F  K  G  K  Y  A  F  L  K

               CTTTGATTGAAGCTATTCAAAAGAACGCTACTGCTTTCAAGGGTAAGTACGCTTTCTTGA

          301 ---------+---------+---------+---------+---------+---------+ 360

    5          GAAACTAACTTCGATAAGTTTTCTTGCGATGACGAAAGTTCCCATTCATGCGAAAGAACT

                                      CP-6
                                           CP-7

                T  Y  N  Y  T  L  G  A  D  D  L  T  P  F  G  E  N  Q  M  V

               AGACTTACAACTACACTTTGGGTGCTGACGACTTGACTCCATTCGGTGAAAACCAAATGG

   10      361 ---------+---------+---------+---------+---------+---------+ 420

               TCTGAATGTTGATGTGAAACCCACGACTGCTGAACTGAGGTAAGCCACTTTTGGTTTACC

                 N  S  G  I  K  F  Y  R  R  Y  K  A  L  A  R  K  I  V  P  F

               TTAACTCTGGTATTAAGTTCTACAGAAGATACAAGGCTTTGGCTAGAAAGATTGTTCCAT

   15      421 ---------+---------+---------+---------+---------+---------+ 480

               AATTGAGACCATAATTCAAGATGTCTTCTATGTTCCGAAACCGATCTTTCTAACAAGGTA

                                                CP-8.7

                                                     CP-9

                I  R  A  S  G  S  S  R  V  I  A  S  A  E  K  F  I  E  G  F

   20          TCATTAGAGCTTCTGGTTCTtctAGAGTTATTGCTTCTGCTGAAAAGTTCATTGAAGGTT
          481 ---------+---------+---------+---------+---------+---------+ 540

               AGTAATCTCGAAGACCAAGAagaTCTCAATAACGAAGACGACTTTTCAAGTAACTTCCAA


                Q  S  A  K  L  A  D  P  G  S  Q  P  H  Q  A  S  P  V  I  D

   25          TCCAATCTGCTAAGTTGGCTGACCCAGGTTCTCAACCACACCAAGCTTCTCCAGTTATTG

          541 ---------+---------+---------+---------+---------+---------+ 600

               AGGTTAGACGATTCAACCGACTGGGTCCAAGAGTTGGTGTGGTTCGAAGAGGTCAATAAC

                                                          CP-10.7
                                                               CP-11.7

   30            V  I  I  S  E  A  S  S  Y  N  N  T  L  D  P  G  T  C  T  A

               ACGTTATTATTtctGAcgctTCTtctTACAACAACACTTTGGACccaGGTACTTGTACTG

          601 ---------+---------+---------+---------+---------+---------+ 660
```

TGCAATAATAAagaCTgcgaAGGagaATGTTGTTGTGAAACCTGggtCCATGAACATGAC

```
                F  E  D  S  E  L  A  D  T  V  E  A  N  F  T  A  L  F  A  P

                CTTTCGAAGACTCTGAATTGgctGACactGTTGAAGCTAACTTCACTGCTTTGTTCGCTC

            661 ---------+---------+---------+---------+---------+---------+ 720

                GAAAGCTTCTGAGACTTAACcgaCTGtgaCAACTTCGATTGAAGTGACGAAACAAGCGAG

  5                                                                CP-12.7

                A  I  R  A  R  L  E  A  D  L  P  G  V  T  L  T  D  T  E  V

                CAGCTATTAGAGCTAGATTGGAAGCTGACTTGCCAGGTGTTACTTTGACTGACactgaaG

            721 ---------+---------+---------+---------+---------+---------+ 780

 10             GTCGATAATCTCGATCTAACCTTCGACTGAACGGTCCACAATGAAACTGACTGtgacttC

                CP-13.7

                T  Y  L  M  D  M  C  S  F  E  T  V  A  R  T  S  D  A  T  E

                TTactTACTTGATGGACATGTGTTcttTCGAAACTGTTGCTAGAACTTCTGACGCTACTG

 15         781 ---------+---------+---------+---------+---------+---------+ 840

                AAtgaATGAACTACCTGTACACAagaAAGCTTTGACAACGATCTTGAAGACTGCGATGAC

                L  S  P  F  C  A  L  F  T  H  D  E  W  R  H  Y  D  Y  L  Q

                AATTGTCTCCATTCTGTGCTTTGTTCACTCACGACGAATGGAGAcacTACGACTACTTGC

 20         841 ---------+---------+---------+---------+---------+---------+ 900

                TTAACAGAGGTAAGACACGAAACAAGTGAGTGCTGCTTACCTCTgtgATGCTGATGAACG

                     CP-14.7
                          CP-15.7

                S  L  K  K  Y  Y  G  H  G  A  G  N  P  L  G  P  T  Q  G  V

 25             AATCTTTGaagAAGTACTACGGTcacGGTGCTGGTAACCCATTGGGTCCAactCAAGGTG

            901 ---------+---------+---------+---------+---------+---------+ 960

                TTAGAAACttcTTCATGATGCCAgtgCCACGACCATTGGGTAACCCAGGTtgaGTTCCAC

                G  F  A  N  E  L  I  A  R  L  T  R  S  P  V  Q  D  H  T  S

 30             TTGGTTTCGCTAACGAATTGATTGCTAGATTGACTAGATCTCCAGTTCAAGACCACACTT

            961 ---------+---------+---------+---------+---------+---------+
     1020

                AACCAAAGCGATTGCTTAACTAACGATCTAACTGATCTAGAGGTCAAGTTCTGGTGTGAA
```

CP-16
CP-17.7

T N H T L D S N P A T F P L N A T L Y A

CTACTAACCACACTTTGGACTCTAACCCAGCTACTTTCCCATTGAACGCTACTTTGTACG

1021 ---------+---------+---------+---------+---------+---------+
1080

GATGATTGGTGTGAAACCTGAGATTGGGTCGATGAAAGGGTAACTTGCGATGAAACATGC

D F S H D N G I I S I F F A L G L Y N G

CTGACTTCTCTCACGACAACggtattATTTCTATTTTCTTCGCTTTGGGTTTGTACAACG

1081 ---------+---------+---------+---------+---------+---------+
1140

GACTGAAGAGAGTGCTGTTGccataaTAAAGATAAAAGAAGCGAAACCCAAACATGTTGC

CP-18.7
CP-19.7

T A P L S T T S V E S I E E T D G Y S S

GTACTGCTCCATTGTCTACTACTTCTGTTGAATCTATTGAAGAAACTGACGGTTACTCTt

1141 ---------+---------+---------+---------+---------+---------+
1200

CATGACGAGGTAACAGATGATGAAGACAACTTAGATAACTTCTTTGACTGCCAATGAGAa

A W T V P F A S R A Y V E M M Q C Q A E

ctgctTGGACTGTTCCATTCgcttctAGAGCTTACGTTGAAATGATGCAATGTCAAGCTG

1201 ---------+---------+---------+---------+---------+---------+
1260

gacgaACCTGACAAGGTAAGcgaagaTCTCGAATGCAACTTTACTACGTTACAGTTCGAC

CP-20
CP-21

K E P L V R V L V N D R V V P L H G C A

AAAAGGAACCATTGGTTAGAGTTTTGGTTAACGACAGAGTTGTTCCATTGCACGGTTGTG

1261 ---------+---------+---------+---------+---------+---------+
1320

TTTTCCTTGGTAACCAATCTCAAAACCAATTGCTGTCTCAACAAGGTAACGTGCCAACAC

V D K L G R C K R D D F V E G L S F A R

CTGTTGACAAGTTGGGTAGATGTAAGAGAGACGACTTCGTTGAAGGTTTGTCTTTCGCTA

1321 ---------+---------+---------+---------+---------+---------+
1380

5 GACAACTGTTCAACCCATCTACATTCTCTCTGCTGAAGCAACTTCCAAACAGAAAGCGAT

CP-22

S G G N W A E C F A * *Eco* RI

GATCTGGTGGTAACTGGGCTGAATGTTTCGCTTAAGAATTCATATA

1381 ---------+---------+---------+---------+------ 1426

10 CTAGACCACCATTGACCCGACTTACAAAGCGAATTCTTAAGTATAT

Figure 11

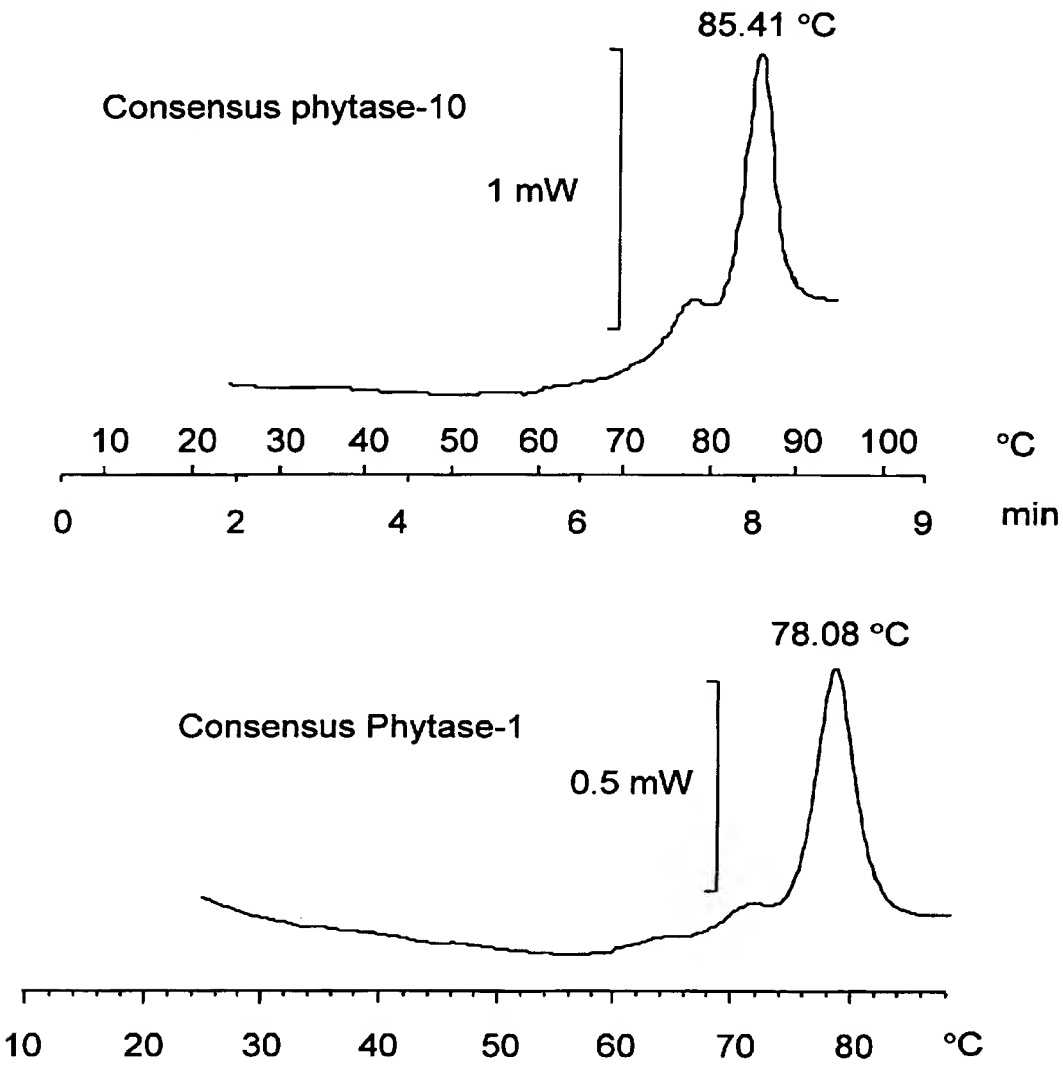


Figure 12

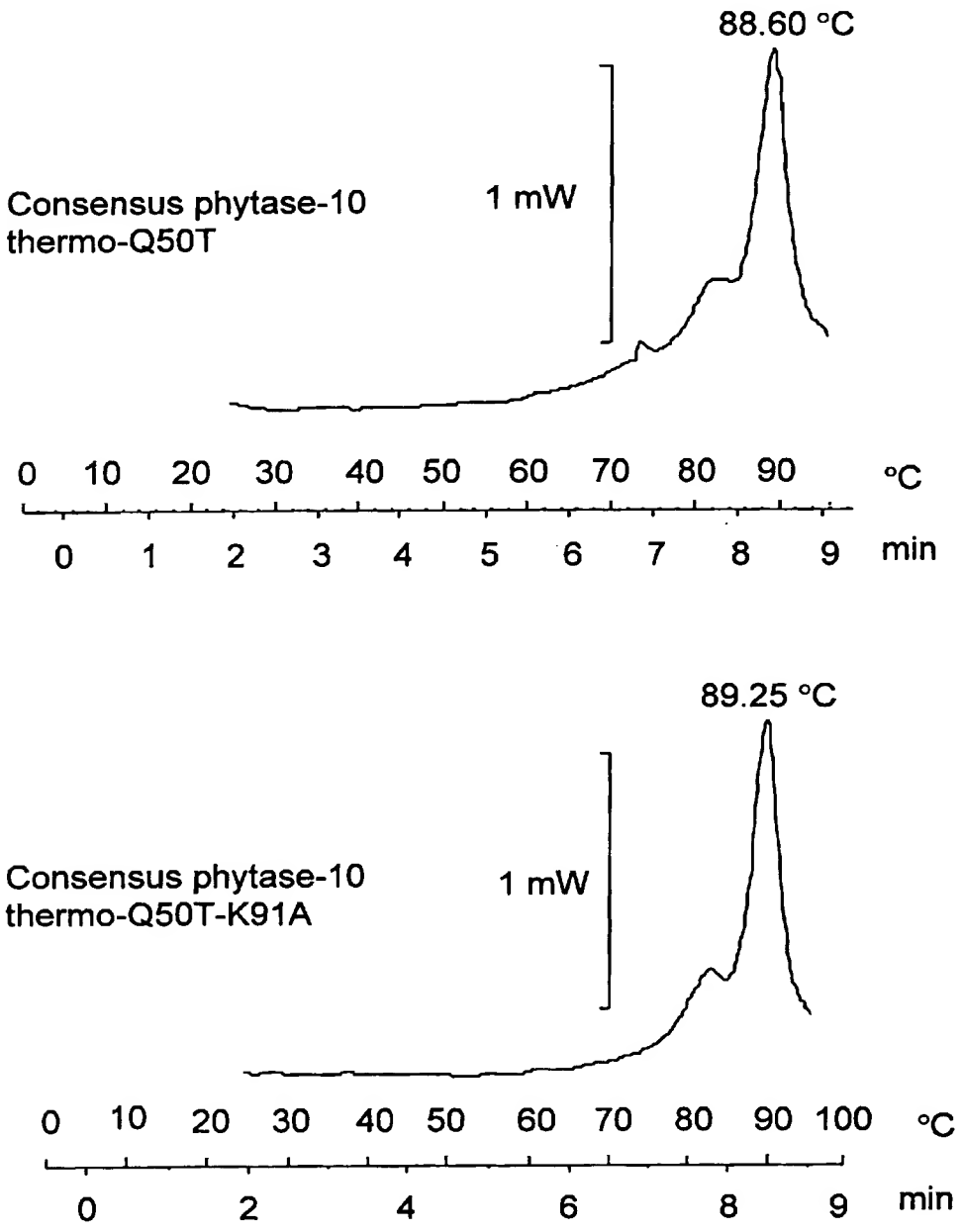


Figure 13

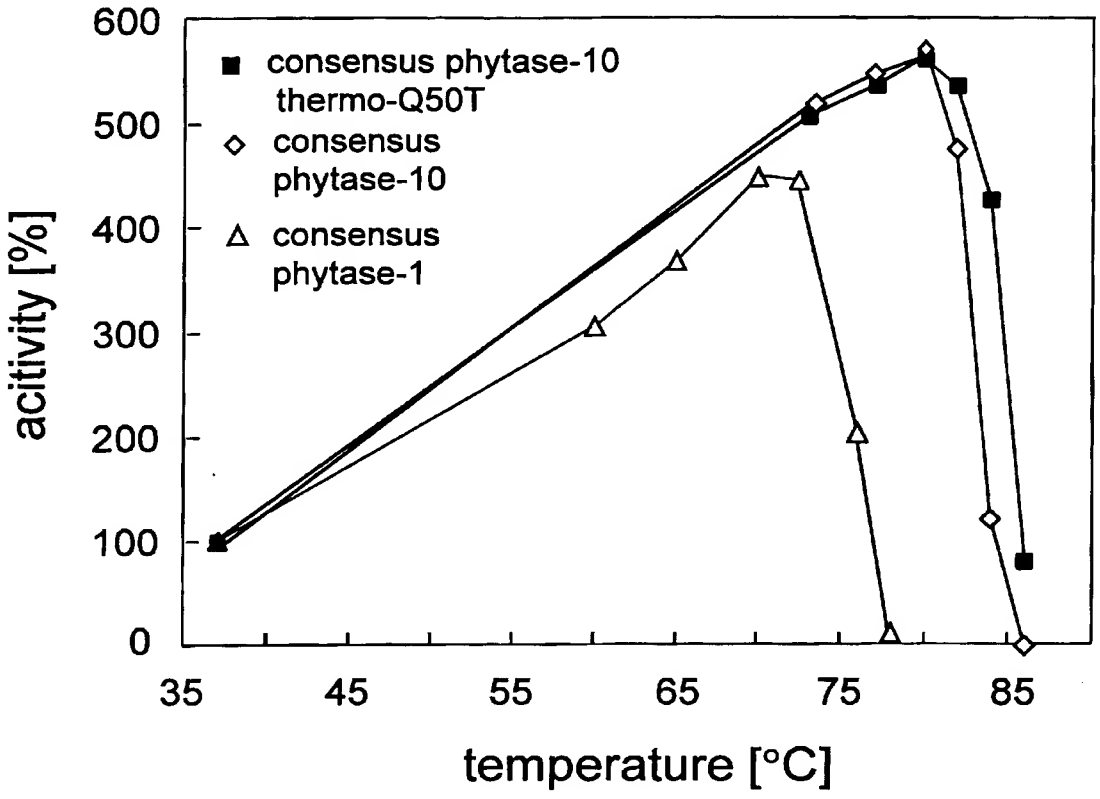


Figure 14

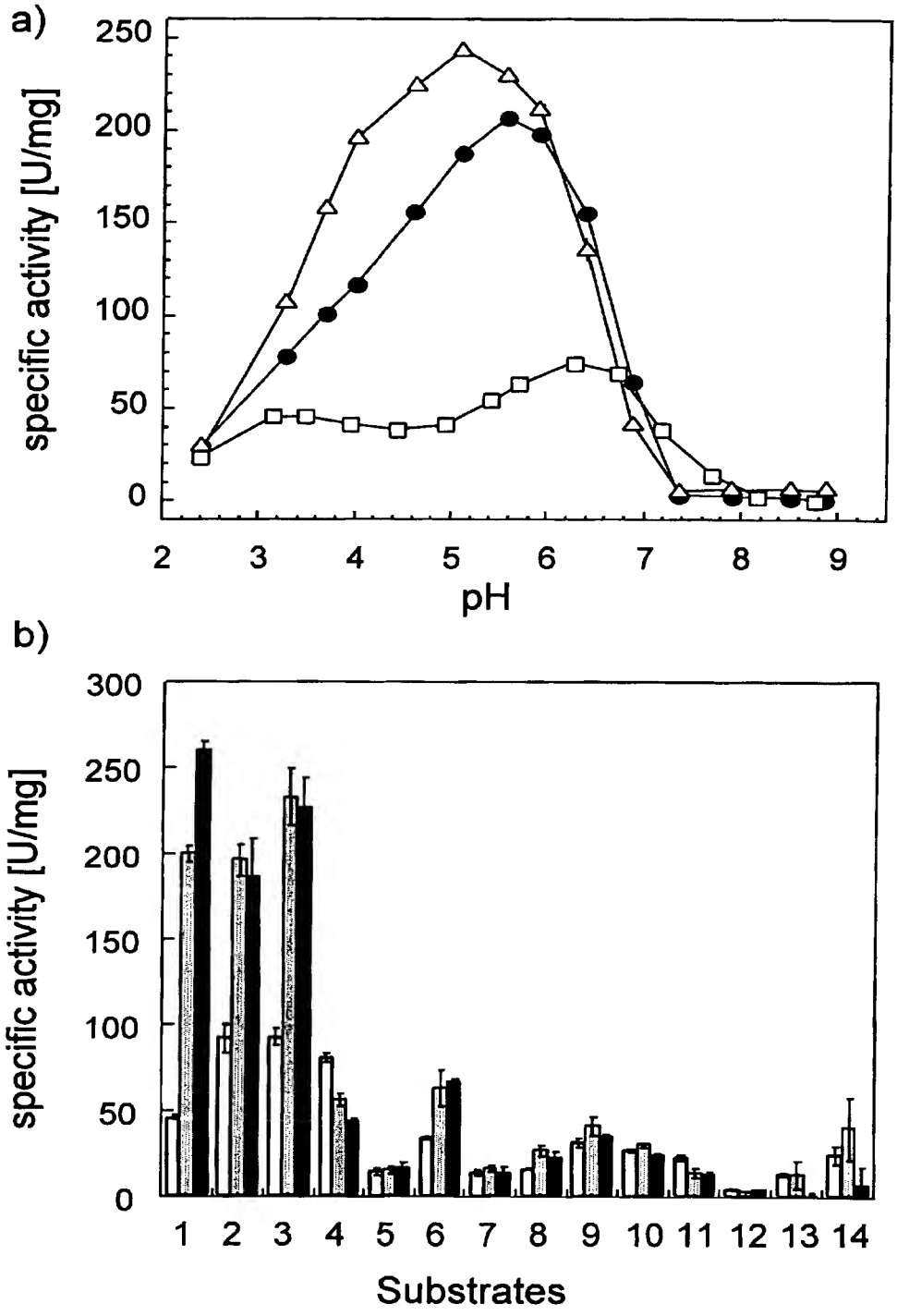
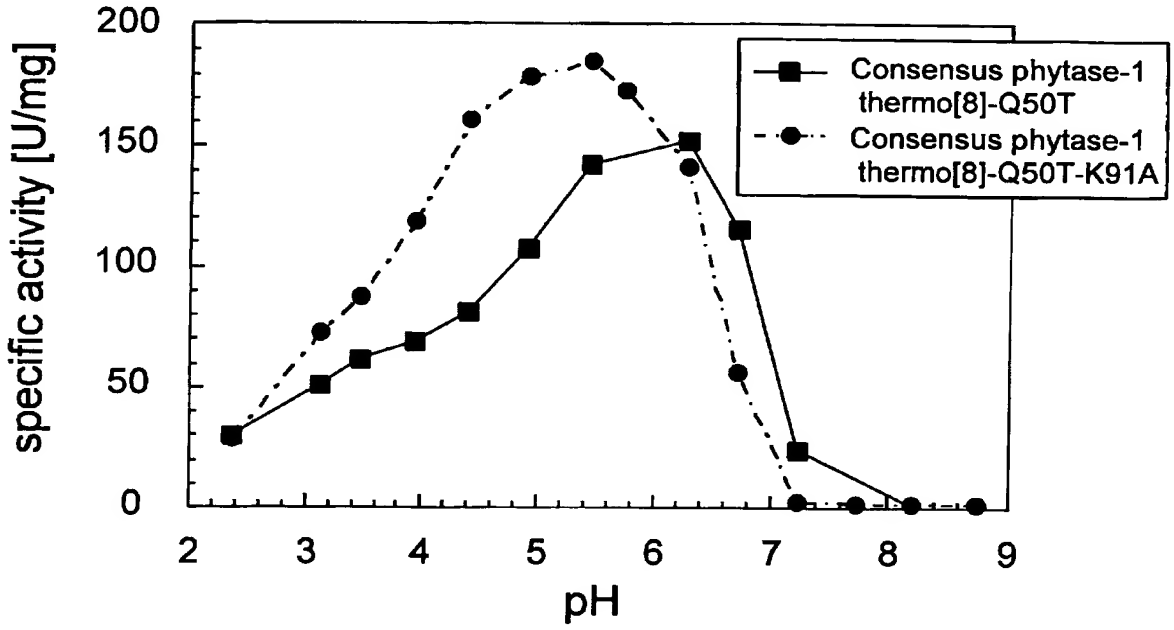


Figure 15



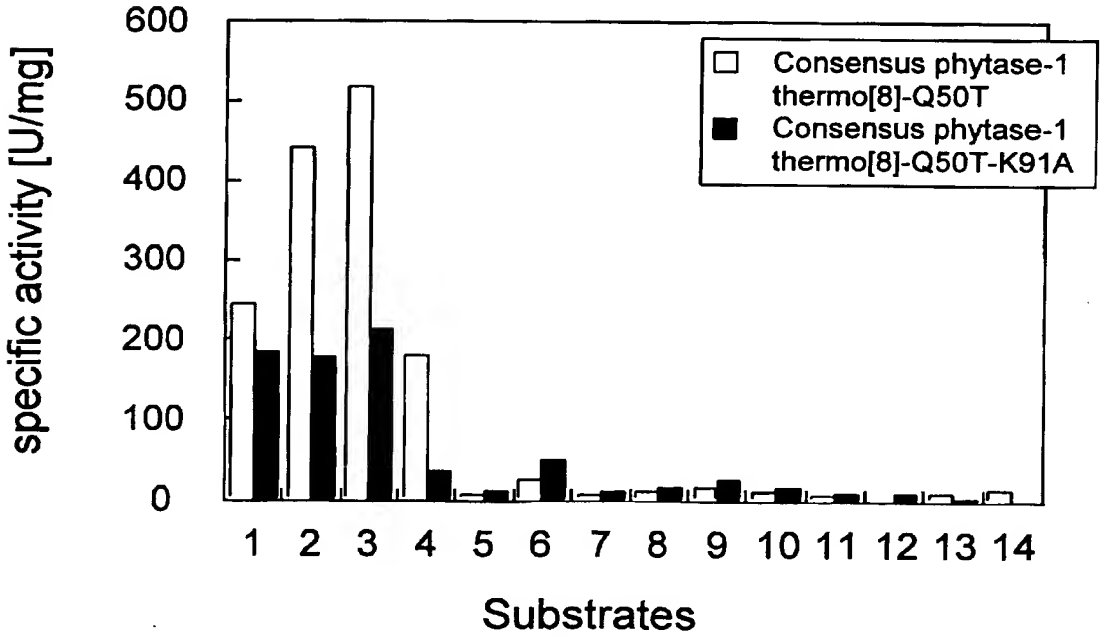


Figure 16

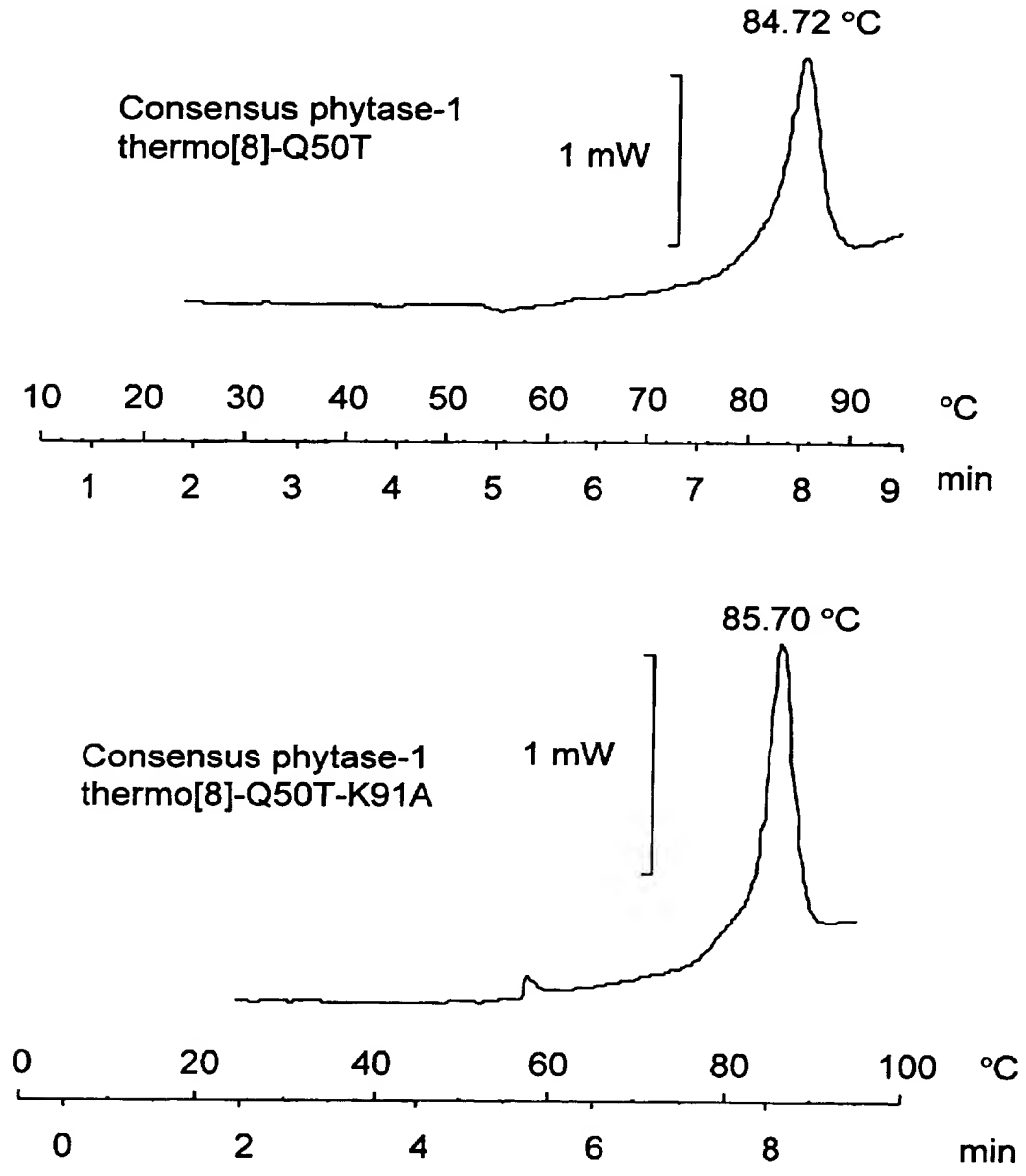


Figure 17

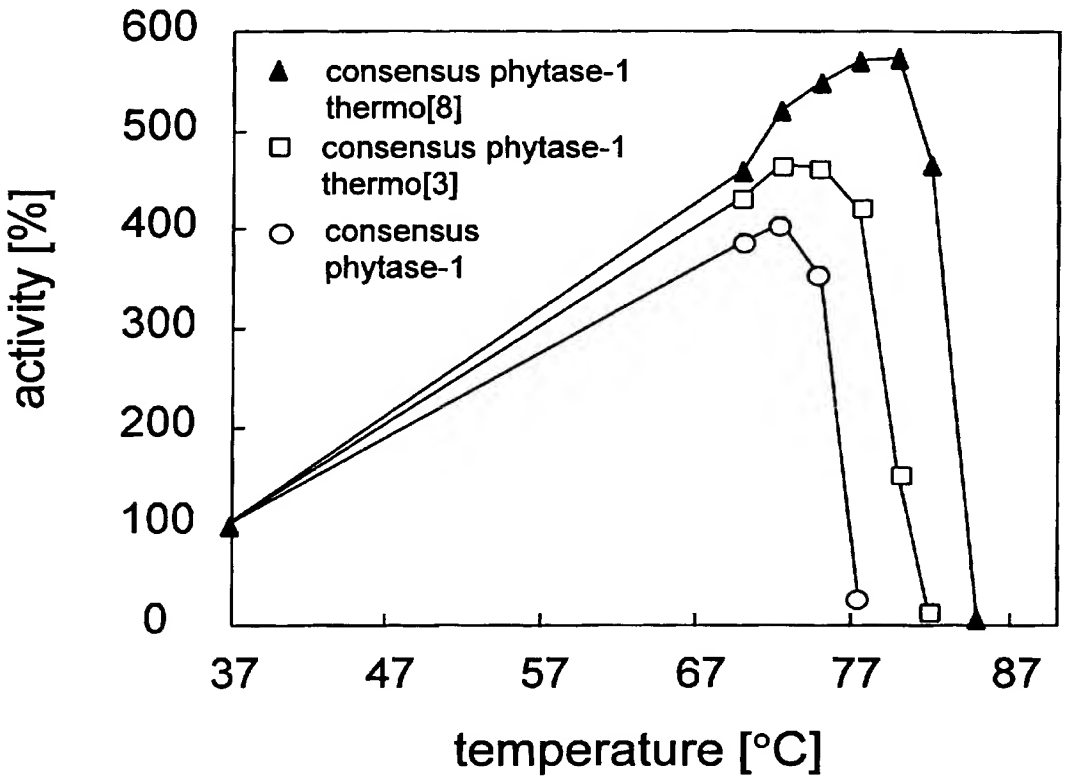
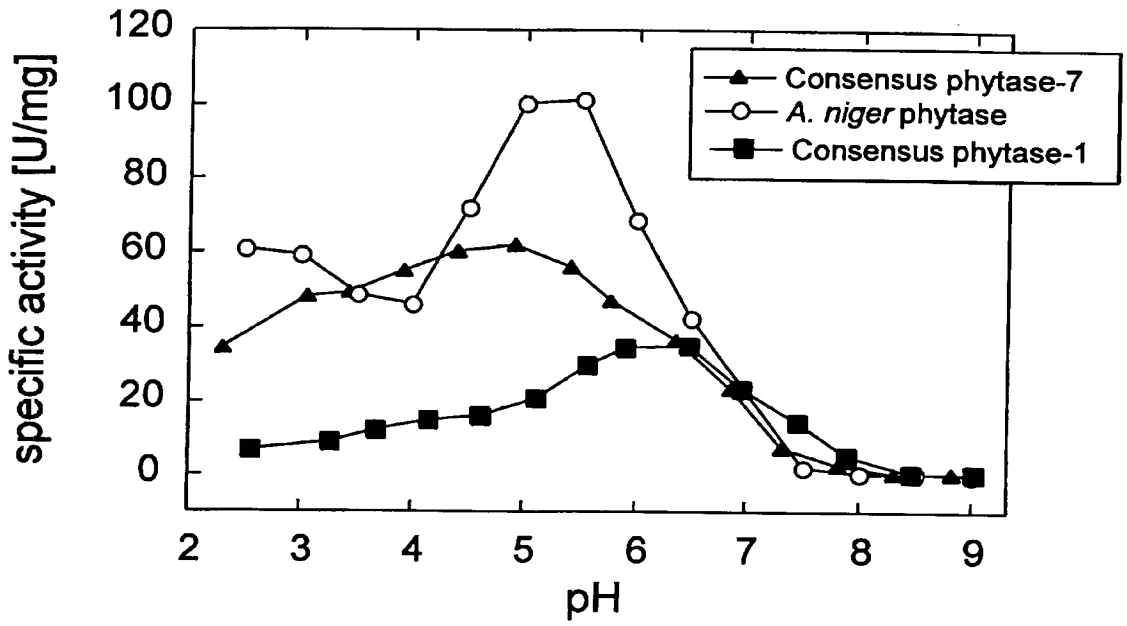


Figure 18



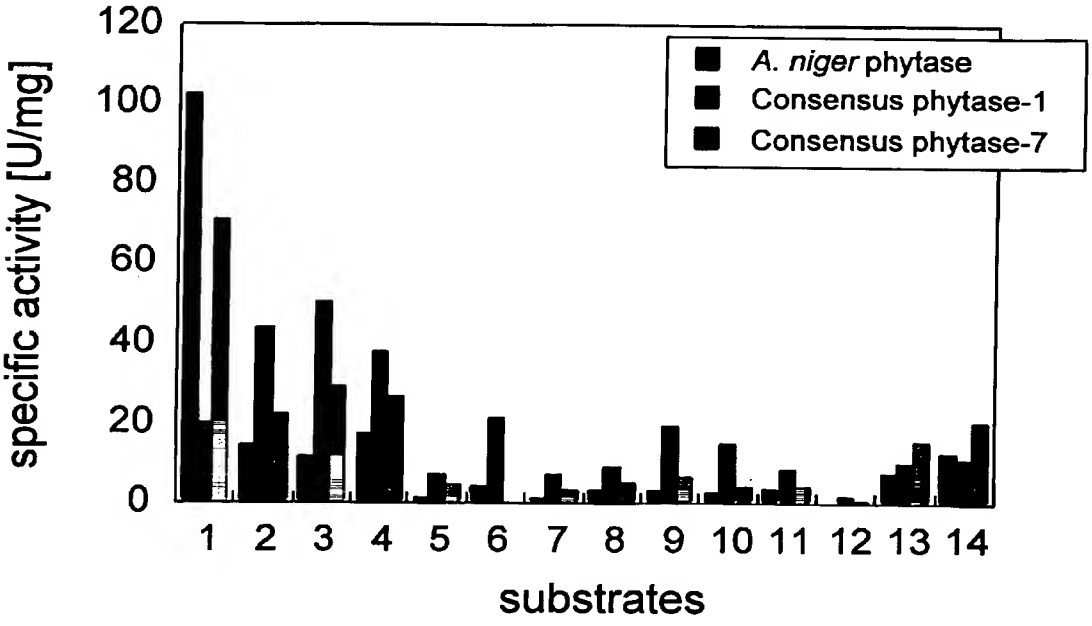


Figure 19

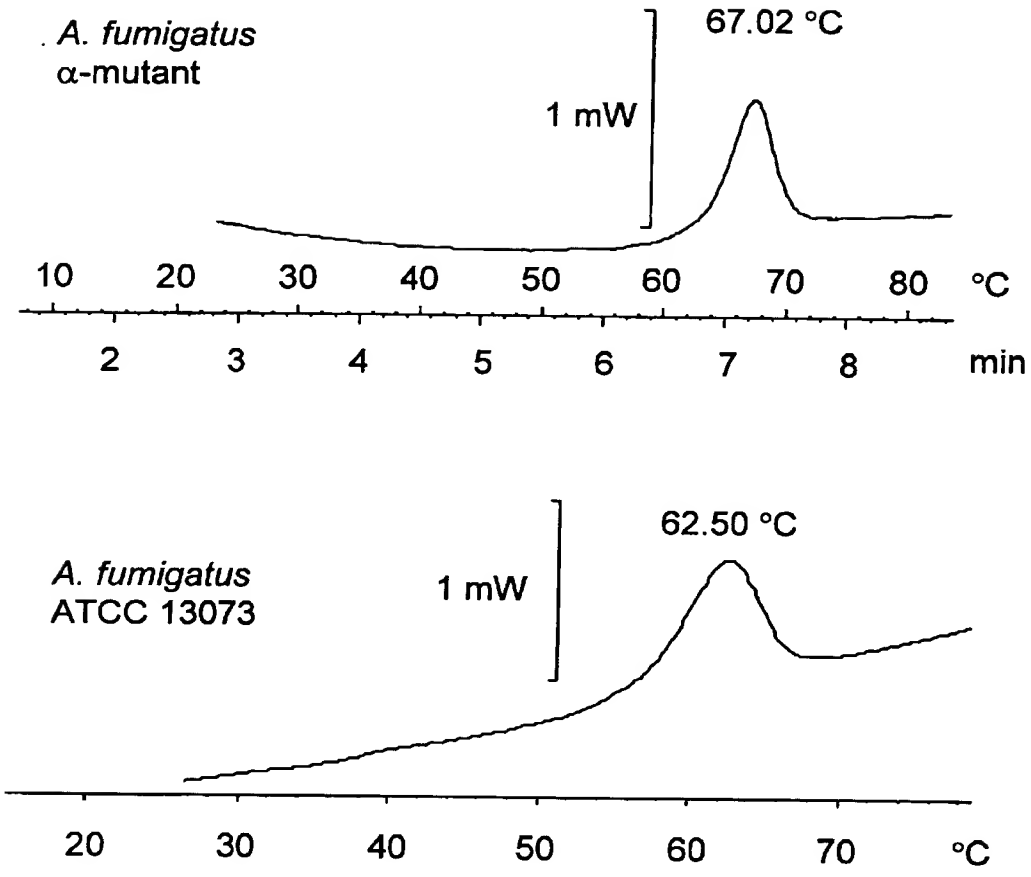


Figure 20

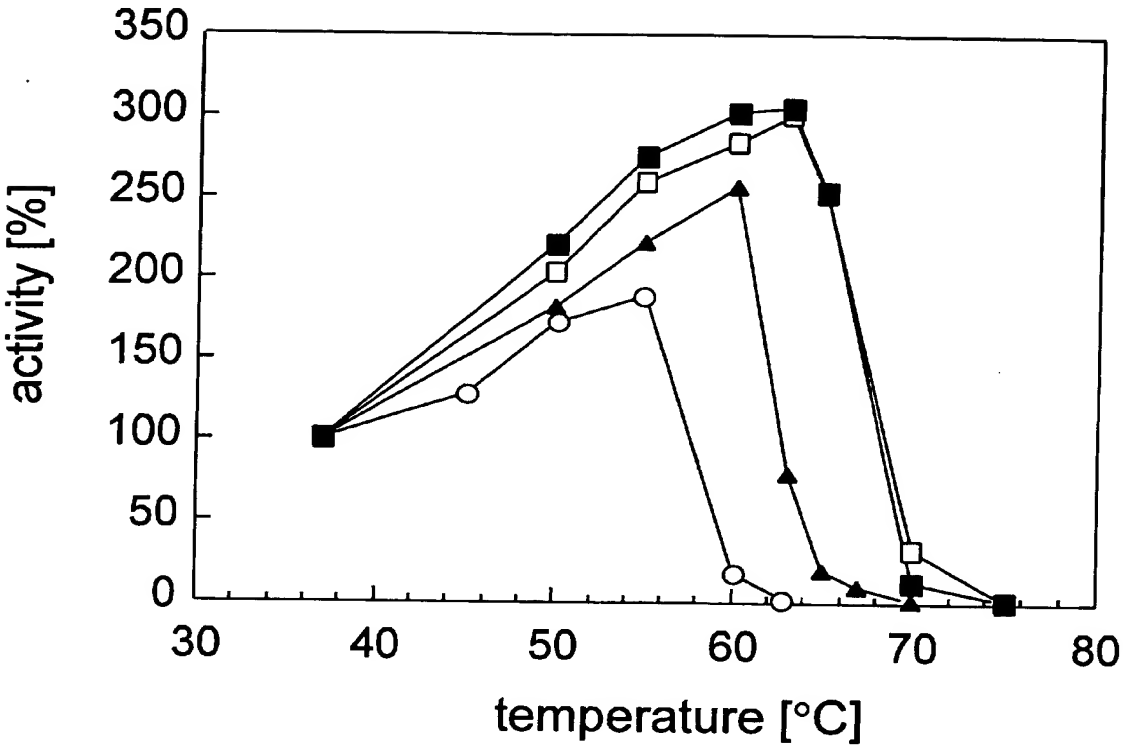


Figure 21

1 MGVFVVLLSI ATLFGSTSGT ALGPRGNSHS CDTVDGGYQC FPEISSNWSP

51 YSPYFSLADE SAISPDVPKG CRVTFVQVLQ RHGARFPTSG AATRISALIE

101 AIQKNATAFK GKYAFLKTYN YTLGADDLVP FGANQSSQAG IKFYRRYKAL

151 ARKIVPFIRA SGSDRVIDSA TNWIEGFQSA KLADPGANPH QASPVINVII

201 PEGAGYNNTL DHGLCTAFEE SELGDDVEAN FTAVFAPPIR ARLEAHLPGV

251 NLTDEDVVNL MDMCPFDTVA RTSDATELSP FCDLFTHDEW IQYDYLGDLD

301 KYYGTGAGNP LGPAQGVGFV NELIARLTHS PVQDHTSTNH TLDSNPATFP

351 LNATLYADFS HDNTMVAIFF ALGLYNGTKP LSTTSVESIE ETDGYSASWL

401 VPFSARMYVE MMQCEAEKEP LVRVLVNDRV VPLHGCGVDK LGRCKRDDFV

451 EGLSFARSGG NWEECFA

Abstract

This invention relates to a new improved consensus phytase by introduction of additional phytase sequences into the sequence alignment and the method of the introduction process. Furthermore, the invention relates to the transfer of stabilizing amino acid exchanges found by the new method into homologous proteins. Furthermore, the invention relates to the replacement of a whole active site of a phytase. It also relates to the corresponding DNA sequences and its generation, methods to produce such phytases and the use thereof.

***